Quantitative Economics 8 (2017), 729–760 1759-7331/20170729

Determining the number of groups in latent panel structures with an application to income and democracy

Xun Lu
Department of Economics, Hong Kong University of Science and Technology

Liangjun Su
School of Economics, Singapore Management University

We consider a latent group panel structure as recently studied by Su, Shi, and Phillips (2016), where the number of groups is unknown and has to be determined empirically. We propose a testing procedure to determine the number of groups. Our test is a residual-based Lagrange multiplier-type test. We show that after being appropriately standardized, our test is asymptotically normally distributed under the null hypothesis of a given number of groups and has the power to detect deviations from the null. Monte Carlo simulations show that our test performs remarkably well in finite samples. We apply our method to study the effect of income on democracy and find strong evidence of heterogeneity in the slope coefficients. Our testing procedure determines three latent groups among 74 countries.

Keywords. Classifier Lasso, dynamic panel, latent structure, penalized least square, number of groups, test.

JEL classification. C12, C23, C33, C38, C52.

1. Introduction

Recently, latent group structures have received much attention in the panel data literature; see, for example, Sun (2005), Lin and Ng (2012), Deb and Trivedi (2013), Bonhomme and Manresa (2015; BM hereafter), Sarafidis and Weber (2015), Ando and Bai (2016), Bester and Hansen (2016), Su, Shi, and Phillips (2016; SSP hereafter), and Su and Ju (forthcoming). In comparison with some other popular approaches to model unobserved heterogeneity in panel data models such as random coefficient models (see, e.g., Hsiao (2014, Chapter 6)), one important advantage of the latent group structure is that it allows flexible forms of unobservable heterogeneity while remaining parsimonious.

Xun Lu: xunlu@ust.hk
Liangjun Su: ljsu@smu.edu.sg
The authors express their sincere appreciation to a co-editor and three anonymous referees for their many constructive comments on the previous versions of the paper. They also thank Stéphane Bonhomme, Xiaohong Chen, Han Hong, Cheng Hsiao, Hidehiko Ichimura, Peter C. B. Phillips, and Katsumi Shimotsu for discussions on the subject matter and valuable comments on the paper. Lu acknowledges support from the Hong Kong Research Grants Council (RGC) under Grant 699513. Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under Grant MOE2012-T2-2-021 and the funding support provided by the Lee Kong Chian Fund for Excellence. All errors are the authors' sole responsibilities.

Copyright © 2017 The Authors. Quantitative Economics. The Econometric Society. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at http://www.qeconomics.org.
DOI: 10.3982/QE517

In addition, the group structure has sound theoretical foundations from game theory or macroeconomic models where multiplicity of Nash equilibria is expected (cf. Hahn and Moon (2010)). The key question in latent group structures is how to identify each individual's group membership. Bester and Hansen (2016) assume that membership is known and determined by external information, say, external classification or geographic location, while others assume that it is unrestricted and unknown, and propose statistical methods to achieve classification. Sun (2005) uses a parametric multinomial logit regression to model membership. Lin and Ng (2012), BM, Sarafidis and Weber (2015), and Ando and Bai (2016) extend K-means classification algorithms to the panel regression framework. Deb and Trivedi (2013) propose expectation–maximization (EM) algorithms to estimate finite mixture panel data models with fixed effects. Motivated by the sparse feature of the individual regression coefficients under latent group structures, SSP propose a novel variant of the Lasso (least absolute shrinkage and selection operator) procedure, that is, the classifier Lasso (C-Lasso), to achieve classification. While these methods make important contributions by empirically grouping individuals, to implement these methods, we often need to determine the number of groups first. Some information criteria have been proposed to achieve this goal (see, e.g., BM and SSP), which often rely on certain tuning parameters. This paper provides a hypothesis-testing-based solution to determine the number of groups.

Specifically, we consider the panel data structure as in SSP,

$$y_{it} = \beta_i^{0\prime} X_{it} + \mu_i + u_{it}, \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \tag{1.1}$$

where X_{it}, μ_i, and u_{it} are the vector of regressors, individual fixed effect, and idiosyncratic error term, respectively, and β_i^0 is the slope coefficient that can depend on individual i. We assume that the N individuals belong to K groups and all individuals in the same group share the same slope coefficients. That is, β_i^0s are homogeneous within each of the K groups but heterogeneous across the K groups. For a given K, we can apply SSP's C-Lasso procedure or the K-means algorithm to determine the group membership and to estimate β_i^0s. However, in practice, K is unknown and has to be determined from data. This motivates us to test the hypothesis

$$\mathbb{H}_0(K_0) : K = K_0 \quad \text{versus} \quad \mathbb{H}_1(K_0) : K_0 < K \leq K_{\max},$$

where K_0 and $K_{\max}$ are prespecified by researchers. We can sequentially test the hypotheses $\mathbb{H}_0(1), \mathbb{H}_0(2), \ldots,$ until we fail to reject $\mathbb{H}_0(K^*)$ for some $K^* \leq K_{\max}$ and conclude that the number of groups is K^*. Onatski (2009) applies a similar procedure to determine the number of latent factors in panel factor structures.

In addition to helping to determine the number of groups, testing $\mathbb{H}_0(K_0)$ itself is also useful for empirical research. When $K_0 = 1$, the test becomes a test for homogeneity in the slope coefficients, which is often assumed in empirical applications. When K_0 is some integer greater than 1, we test whether the group structure is correctly specified. Although the group structure is flexible in terms of modeling unobserved slope heterogeneity, it could still be misspecified. Inferences based on misspecified models are often misleading. Thus conducting a formal specification test is highly desirable. Although we work in the same framework as SSP, the questions of interest are different. SSP study the classification and estimation problem, while we consider the specification testing. To avoid overlapping with SSP's paper, we omit the detailed discussions on the estimation in this paper.

Our test is a residual-based Lagrange multiplier (LM) type test. We estimate the model under the null hypothesis $\mathbb{H}_0(K_0)$ to obtain the restricted residuals, and the test statistic is based on whether the regressors have predictive power for the restricted residuals. Under the null of the correct number of latent groups, the regressors should not contain any useful information about the restricted residuals. We show that after being appropriately standardized, our test statistic is asymptotically normal under the null. The p values can be obtained based on the standard normal approximation, and thus the test is easy to implement. Our test is related to the literature on testing slope homogeneity and poolability for panel data models in which $K_0 = 1$. See, for example, Pesaran, Smith, and Im (1996), Phillips and Sul (2003), Pesaran and Yamagata (2008), and Su and Chen (2013), among others. Nevertheless, none of the existing tests can be directly applied to test $K = K_0$, where $K_0 > 1$. Also, the test proposed in this paper substantially differs from the existing tests in technical details, as we need to apply the C-Lasso method or K-means algorithm to estimate the model under the null.

When there are no time fixed effects in the model, both SSP's C-Lasso method and the K-means algorithm can be used to estimate the model under $\mathbb{H}_0(K_0)$. So the residuals that are used to construct our LM statistic can be obtained from either method. When time fixed effects are present in the model, we extend SSP's method to allow for time fixed effects and show that the LM statistic behavior is asymptotically equivalent to that in the absence of time fixed effects.

We conduct Monte Carlo simulations to show the excellent finite-sample performance of our test. Both the levels and powers of our test perform well in finite samples. For the data generating processes (DGPs) considered, when both N and T are large, our method can determine the number of groups correctly with a high probability.

Our method is applicable to a wide range of empirical studies. In this paper, we provide detailed empirical analysis on the relationship between income and democracy. Specifically, y_{it} is a measure of democracy for country i in period t: X_{it} includes its income (the logarithm of its real gross domestic product (GDP) per capita) and lagged y_{it}. The main parameter of interest is the effect of income on democracy. In an influential paper, Acemoglu et al. (2008; AJRY hereafter) use the standard panel data model with common slope coefficients and find that the effect of income on democracy is insignificant. We find that the slope coefficients (the effects of income and lagged democracy on democracy) are actually heterogeneous. The hypothesis of homogeneous slope coefficients is strongly rejected with p values being less than 0.001. This suggests that the common coefficient model is likely to be misspecified. Further, we determine the number of heterogeneous groups to be three and find that the slope coefficients of the three groups differ substantially. In particular, for one group, the effect of income on democracy is positive and significant, and for the other two groups, the income effects are also significant, but negative with different magnitudes. Therefore, our method provides the new insight that the effect of income on democracy is heterogeneous and significant, in sharp contrast to the existing finding that the income effect is insignificant. Our classification of groups is completely data-driven and the result does not show any apparent pattern, for example, in geographic locations. We further investigate the determinants of the group pattern by running a cross-country multinomial logit regression on some country-specific covariates. We find that two variables, namely, the constraints on the executive at independence and the long-run economic growth, are important determinants.

There are numerous potential applications of our method. For example, SSP study the determinants of saving rates across countries. That is, Y_{it} is the ratio of saving to GDP, and X_{it} includes the lagged saving rates, inflation rates, real interest rates, and per capita GDP growth rates. Using an information criterion, they find a two-group structure in the slope coefficients. Our method is perfectly applicable to their setting. Another example is Hsiao and Tahmiscioglu (1997, HT), who study the effect of firms' liquidity constraints on their investment expenditure. Specifically, their Y_{it} is the ratio of firm i's capital investment to its capital stock at time t and X_{it} includes its liquidity (defined as cash flow minus dividends), sales, and Tobin's q. HT classify firms into two groups—less-capital-intensive firms and more-capital-intensive firms—using the capital intensity ratio of 0.55 as a cutoff point. Our method can be applied to their setting by classifying the firms into heterogeneous groups in a data-driven way, which allows unobservable heterogeneity in the slope coefficients. In general, our method provides a powerful tool to detect unobserved heterogeneity, which is an important phenomenon in panel data analysis. It can be applied to any linear model for panel data where the time dimension is relatively long.

There are two limitations of our approach. First, we treat the unknown parameters as fixed; thus our asymptotic analysis is only pointwise. We do not consider the uniform inference here. Second, our method requires that both N and T be large. In particular, the finite-sample results may not be reliable when T is small, and the poor performance can occur when the true number of groups is large relative to N or the differences between the values of the parameters in different groups are small.

The remainder of the paper is organized as follows. In Section 2, we introduce the hypotheses and the test statistic for panel data models with individual fixed effects only. In Section 3, we derive the asymptotic distribution of our test statistic under the null and study the global power of our test. In Section 4, we extend the analysis to allow for both individual and time fixed effects in the models. We conduct Monte Carlo experiments to evaluate the finite-sample performance of our test in Section 5 and apply it to the income–democracy data set in Section 6. Section 7 concludes. All proofs are relegated to the Appendix, available in a supplementary file on the journal website, http://qeconomics.org/supp/517/supplement.pdf.

Notation. For an $m \times n$ real matrix A, we denote its transpose as A' and its Frobenius norm as $\|A\|$ ($\equiv [\mathrm{tr}(AA')]^{1/2}$), where $\equiv$ means "is defined as." Let $P_A \equiv A(A'A)^{-1}A'$ and $M_A \equiv I_m - P_A$, where I_m denotes an $m \times m$ identity matrix. When $A = \{a_{ij}\}$ is symmetric, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote its maximum and minimum eigenvalues, respectively, and denote $\mathrm{diag}(A)$ as a diagonal matrix whose (i,i)th diagonal element is given by a_{ii}. Let $P_0 \equiv T^{-1}\mathbf{i}_T\mathbf{i}_T'$ and $M_0 \equiv I_T - T^{-1}\mathbf{i}_T\mathbf{i}_T'$, where $\mathbf{i}_T$ is a $T \times 1$ vector of

1s. Moreover, the operator $\longrightarrow^{P}$ denotes convergence in probability, and $\longrightarrow^{D}$ denotes convergence in distribution. We use $(N, T) \to \infty$ to denote the joint convergence of N and T when N and T pass to infinity simultaneously. We abbreviate positive semidefinite, with probability approaching 1, and without loss of generality as p.s.d., w.p.a.1, and wlog, respectively.

2. Hypotheses and test statistic

In this section, we introduce the hypotheses and test statistic.

2.1 *Hypotheses*

We consider the panel structure model

$$y_{it} = \beta_i^{0\prime} X_{it} + \mu_i + u_{it}, \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \tag{2.1}$$

where X_{it} is a $p \times 1$ vector of strictly exogenous or predetermined regressors, μ_i is an individual fixed effect, and u_{it} is the idiosyncratic error term. The model with both individual and time fixed effects will be studied in Section 4. We assume that β_i^0 has the group structure

$$\beta_i^0 = \sum_{k=1}^{K_0} \alpha_k^0 \mathbf{1}\{i \in G_k^0\}, \tag{2.2}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function, K is an integer, and $\{G_1^0, \ldots, G_K^0\}$ forms a partition of $\{1, \ldots, N\}$ such that $\bigcup_{k=1}^{K} G_k^0 = \{1, \ldots, N\}$ and $G_k^0 \cap G_j^0 = \emptyset$ for any $j \neq k$. Further, $\alpha_k^0 \neq \alpha_j^0$ for any $j \neq k$. Let $N_k = \# G_k^0$ denote the cardinality of the set G_k^0. We assume that K, $\mathcal{G}^0 \equiv \{G_1^0, \ldots, G_K^0\}$, $\boldsymbol{\alpha}_K^0 \equiv (\alpha_1^0, \ldots, \alpha_K^0)$, and $\boldsymbol{\beta}^0 \equiv (\beta_1^0, \ldots, \beta_N^0)$ are all unknown. One key step in estimating all these parameters is to first determine K, as once K is determined, we can readily apply SSP's C-Lasso method or the K-means algorithm. This motivates us to test the hypothesis

$$\mathbb{H}_0(K_0) : K = K_0 \quad \text{versus} \quad \mathbb{H}_1(K_0) : K_0 < K \leq K_{\max}. \tag{2.3}$$

Here we assume that K, $K_{\max}$, and p are fixed, that is, they do not increase with the sample size N or T.

The testing procedure developed below can be used to determine K. Suppose that we have a priori information such that $K_{\min} \leq K \leq K_{\max}$, where $K_{\min}$ is typically 1. Then we can first test $\mathbb{H}_0(K_{\min})$ against $\mathbb{H}_1(K_{\min})$. If we fail to reject the null, then we conclude that $K = K_{\min}$. Otherwise, we continue to test $\mathbb{H}_0(K_{\min} + 1)$ against $\mathbb{H}_1(K_{\min} + 1)$. We repeat this procedure until we fail to reject the null $\mathbb{H}_0(K^*)$ and conclude that $K = K^*$. If we reject $K = K_{\max}$, then we can use the random coefficient model to estimate the model, that is, $K = N$. This procedure is also used in other contexts, for example, to determine the number of lags in autoregressive (AR) models, the cointegration rank, the rank of a matrix, and the number of latent factors in panel factor structures Onatski (2009).

In theory, $K_{\max}$ can be any finite number. However, in practice, we do not suggest choosing too large a value for $K_{\max}$. Our method is powerful when the number of groups is small. When the number of groups is large, the classification errors can be large. Therefore, if we reject $K = K_{\max}$ for a reasonably large $K_{\max}$, then we can simply adopt the random coefficient model to avoid classification errors.

2.2 *Estimation under the null and test statistic*

Our test is a residual-based test and so we only need to estimate the model under $\mathbb{H}_0(K_0) : K = K_0$. In the special case where $K_0 = 1$, the panel structure model reduces to a homogeneous panel data model so that $\beta_i^0 = \beta^0$ for all $i = 1, \dots, N$, and we can estimate the homogeneous slope coefficient using the usual within-group estimator $\hat{\beta}$.

In the general case where $K_0 > 1$, we can consider two popular ways to estimate the unknown group structure and the group-specific parameters. For example, we can apply SSP's C-Lasso procedure. Let $\tilde{y}_{it} = y_{it} - \bar{y}_{i\cdot}$, where $\bar{y}_{i\cdot} = T^{-1} \sum_{t=1}^{T} y_{it}$. Define $\tilde{X}_{it}$ and $\bar{X}_{i\cdot}$ analogously. Let $\tilde{\boldsymbol{\beta}} \equiv (\tilde{\beta}_1, \dots, \tilde{\beta}_N)$ and $\tilde{\boldsymbol{\alpha}}_{K_0} \equiv (\tilde{\alpha}_1, \dots, \tilde{\alpha}_{K_0})$ be the C-Lasso estimators proposed in SSP, which are defined as the minimizer of the criterion function

$$Q_{1NT,\lambda}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}_{K_0}) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda}{N} \sum_{i=1}^{N} \prod_{k=1}^{K_0} \big\| \beta_i - \alpha_k \big\|, \tag{2.4}$$

where $\lambda \equiv \lambda_{NT}$ is a tuning parameter and

$$Q_{1,NT}(\boldsymbol{\beta}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \big(\tilde{y}_{it} - \beta_i' \tilde{X}_{it}\big)^2.$$

Let $\hat{G}_k = \{i \in \{1, 2, \dots, N\} : \tilde{\beta}_i = \tilde{\alpha}_k\}$ for $k = 1, \dots, K_0$. Let $\hat{G}_0 = \{1, 2, \dots, N\} \setminus (\bigcup_{k=1}^{K_0} \hat{G}_k)$. Although SSP demonstrate that the number of elements in $\hat{G}_0$ shrinks to zero as $T \to \infty$, in finite samples, $\hat{G}_0$ may not be empty. To fully impose the null hypothesis $\mathbb{H}_0(K_0)$, we can force all the estimates of the slope coefficients to be grouped into K_0 groups. Specifically, we classify member i in $\hat{G}_0$ to group k^*, where $k^* \equiv \arg\min_k \{\|\tilde{\beta}_i - \tilde{\alpha}_k\|, k = 1, \dots, K_0\}$. Our final estimators of β_i^0s are the post-Lasso estimators. Specifically, for each of the classified groups, we reestimate the homogeneous slope coefficient using the usual within-group estimator. Let $\hat{\alpha}_k$ $(k = 1, \dots, K_0)$ be the estimator for group k. Then the final estimators of β_i^0s are $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, \dots, \hat{\beta}_N)$, where $\hat{\beta}_i = \hat{\alpha}_k$ for $i \in \hat{G}_k$.

Alternatively, one can apply the K-means algorithm as advocated by Lin and Ng (2012), BM, Sarafidis and Weber (2015), and Ando and Bai (2016). Let $\mathbf{g} = \{g_1, \dots, g_N\}$ denote the group membership such that $g_i \in \{1, \dots, K_0\}$. The K-means algorithm estimates of $\boldsymbol{\alpha}_{K_0}$ and $\mathbf{g}$ can be obtained as the minimizer of the objective function

$$Q_{NT}^{(K_0)}(\boldsymbol{\alpha}_{K_0}, \mathbf{g}) = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i:g_i=k} \sum_{t=1}^{T} \big(\tilde{y}_{it} - \alpha_{g_i}' \tilde{X}_{it}\big)^2. \tag{2.5}$$

Let $\hat{\mathbf{g}} \equiv \{\hat{g}_1, \dots, \hat{g}_N\}$ denote the K-means algorithm estimate of $\mathbf{g}$. With a little abuse of notation, we also use $\hat{\boldsymbol{\alpha}}_{K_0} \equiv (\hat{\alpha}_1, \dots, \hat{\alpha}_{K_0})$ and $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, \dots, \hat{\beta}_N)$ to denote the K-means algorithm estimate of $\boldsymbol{\alpha}^0_{K_0}$ and $\boldsymbol{\beta}^0$, where $\hat{\beta}_i = \hat{\alpha}_{\hat{g}_i}$. Define $\hat{G}_k = \{i \in \{1, 2, \dots, N\} : \hat{g}_i = k\}$ for $k = 1, \dots, K_0$. Note that the K-means algorithm forces all individuals to be classified into one of the K_0 groups automatically so that $\hat{G}_0$ becomes an empty set in this case.

Given $\{\hat{\beta}_i\}$, we can estimate the individual fixed effects using $\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^T (y_{it} - \hat{\beta}_i' X_{it})$.[1] The residuals are obtained by

$$\hat{u}_{it} \equiv y_{it} - \hat{\beta}_i' X_{it} - \hat{\mu}_i. \tag{2.6}$$

It is easy to show that

$$\begin{aligned} \hat{u}_{it} &= (y_{it} - \bar{y}_{i\cdot}) - (X_{it} - \bar{X}_{i\cdot})' \hat{\beta}_i \\ &= u_{it} - \bar{u}_{i\cdot} + (X_{it} - \bar{X}_{i\cdot})' \big(\beta_i^0 - \hat{\beta}_i\big), \end{aligned} \tag{2.7}$$

where $\bar{u}_{i\cdot} = T^{-1}\sum_{t=1}^T u_{it}$. Under the null hypothesis, $\hat{\beta}_i$ is a consistent estimator of β_i^0.[2] Hence, $\hat{u}_{it}$ should be close to u_{it}. By the assumption that the regressors are predetermined, X_{it} should not have any predictive power for u_{it}. Specifically, we assume that the error term u_{it} is a martingale difference sequence (m.d.s.) (Assumption 1(v) below), which implies that $E(u_{it} \mid X_{it}) = 0$. If we write u_{it} in linear regression form, we have

$$u_{it} = v_i + \phi_i' X_{it} + \eta_{it}, \quad i = 1, \dots, N, t = 1, \dots, T,$$

where v_i and ϕ_i are the intercept and slope coefficients, respectively, and η_{it} is the regression error. Then the m.d.s. assumption implies that $\phi_i = 0$ for all i's. This motivates us to run the auxiliary regression model

$$\hat{u}_{it} = v_i + \phi_i' X_{it} + \eta_{it}, \quad i = 1, \dots, N, t = 1, \dots, T, \tag{2.8}$$

and test the null hypothesis

$$\mathbb{H}_0^* : \phi_i = 0 \quad \text{for all } i = 1, \dots, N.$$

We construct an LM-type test statistic by concentrating the intercept v_i out in (2.8). Consider the Gaussian quasi-likelihood function for $\hat{u}_{it}$,

$$\ell(\boldsymbol{\phi}) = \sum_{i=1}^N (\hat{u}_i - M_0 X_i \phi_i)'(\hat{u}_i - M_0 X_i \phi_i),$$

where $\boldsymbol{\phi} \equiv (\phi_1, \dots, \phi_N)'$, $\hat{u}_i \equiv (\hat{u}_{i1}, \dots, \hat{u}_{iT})'$, $X_i \equiv (X_{i1}, \dots, X_{iT})'$, $M_0 \equiv I_T - T^{-1}\mathbf{i}_T \mathbf{i}_T'$, and $\mathbf{i}_T$ is a $T \times 1$ vector of 1s. Define the LM statistic as

$$\mathrm{LM}_{1NT}(K_0) = \left(T^{-1/2}\frac{\partial \ell(0)}{\partial \boldsymbol{\phi}}\right)' \left(-T^{-1}\frac{\partial^2 \ell(0)}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}\right)^{-1} \left(T^{-1/2}\frac{\partial \ell(0)}{\partial \boldsymbol{\phi}}\right), \tag{2.9}$$

[1] If $K_0 = 1$, we set $\hat{\beta}_i = \hat{\beta}$, the within-group estimator of the homogeneous slope coefficient. Note that we also suppress the dependence of $\hat{\mu}_i$ on K_0.

[2] Strictly speaking, $\tilde{\beta}_i$ is a consistent estimator of β_i^0 under the null. But because the cardinality of the set $\hat{G}_0$ shrinks to zero under the null as $T \to \infty$, the difference between $\tilde{\beta}_i$ and $\hat{\beta}_i$ is asymptotically negligible.

where we make the dependence of $\mathrm{LM}_{1NT}(K_0)$ on K_0 explicit. We can verify that

$$\mathrm{LM}_{1NT}(K_0)=\sum_{i=1}^{N}\hat{u}_i'M_0X_i\left(X_i'M_0X_i\right)^{-1}X_i'M_0\hat{u}_i, \tag{2.10}$$

where the dependence of $\mathrm{LM}_{1NT}(K_0)$ on K_0 is through that of $\hat{u}_i$ on K_0. We will show that after being appropriately scaled and centered, $\mathrm{LM}_{1NT}(K_0)$ is asymptotically normally distributed under $\mathbb{H}_0(K_0)$ and diverges to infinity under $\mathbb{H}_1(K_0)$.

Remark 2.1. We have included a constant term in the regression in (2.8). Under the assumption that $E(u_{it})=0$ and N and T pass to infinity jointly, one can also omit the constant term and obtain the LM test statistic

$$\overline{\mathrm{LM}}_{1NT}(K_0)=\sum_{i=1}^{N}\hat{u}_i'X_i\left(X_i'X_i\right)^{-1}X_i'\hat{u}_i. \tag{2.11}$$

The asymptotic distribution of $\overline{\mathrm{LM}}_{NT}(K_0)$ can be similarly studied with little modification. In case T is not very large as in our empirical applications, we recommend including a constant term in the auxiliary regression in (2.8) and thus only focus on the study of $\mathrm{LM}_{1NT}(K_0)$ below.

3. Asymptotic properties

In this section, we first present a set of assumptions that are necessary for asymptotic analyses and then study the asymptotic distributions of $\mathrm{LM}_{1NT}(K_0)$ under both $\mathbb{H}_0(K_0)$ and $\mathbb{H}_1(K_0)$.

3.1 *Assumptions*

Let $\|A\|_q\equiv[E(\|A\|^q)]^{1/q}$ for $q\geq 1$. Let $\hat{\Omega}_i\equiv T^{-1}X_i'M_0X_i$ and $\Omega_i\equiv E(\hat{\Omega}_i)$. Define $\mathcal{F}_{NT,t}\equiv\sigma(\{X_{i,t+1},X_{it},u_{it},X_{i,t-1},u_{i,t-1},\ldots\}_{i=1}^N)$. Let $C<\infty$ be a generic constant that may vary across lines. Following SSP, we define the two types of classification errors

$$\hat{E}_{kNT,i}=\left\{i\notin\hat{G}_k\mid i\in G_k^0\right\}\quad\text{and}\quad\hat{F}_{kNT,i}=\left\{i\notin G_k^0\mid i\in\hat{G}_k\right\}, \tag{3.1}$$

where $i=1,\ldots,N$ and $k=1,\ldots,K_0$. Let $\hat{E}_{kNT}=\bigcup_{i\in G_k^0}\hat{E}_{kNT,i}$ and $\hat{F}_{kNT}=\bigcup_{i\in\hat{G}_k}\hat{F}_{kNT,i}$.

We make the following assumptions.

Assumption 1. (i) *We have* $\max_{1\leq i\leq N}\max_{1\leq t\leq N}\|\zeta_{it}\|_{8+4\sigma}\leq C$ *for some* $\sigma>0$ *for* $\zeta_{it}=X_{it},u_{it}$, *and* $X_{it}u_{it}$.

(ii) *There exist positive constants* $\underline{c}_\Omega$ *and* $\bar{c}_\Omega$ *such that* $\underline{c}_\Omega\leq\min_{1\leq i\leq N}\lambda_{\min}(\Omega_i)\leq\max_{1\leq i\leq N}\lambda_{\max}(\Omega_i)\leq\bar{c}_\Omega$.

(iii) *For each* $i=1,\ldots,N$, $\{(X_{it},u_{it}):t=1,2,\ldots\}$ *is a strong mixing process with mixing coefficients* $\{\alpha_{NT,i}(\cdot)\}$. *The equality* $\alpha(\cdot)\equiv\alpha_{NT}(\cdot)\equiv\max_{1\leq i\leq N}\alpha_{NT,i}(\cdot)$ *satisfies* $\alpha(s)=O_{a.s.}(s^{-\rho})$, *where* $\rho=3(2+\sigma)/\sigma+\epsilon$ *for some* $\epsilon>0$. *In addition, there exist integers* $\tau_0,\tau_*\in(1,T)$ *such that* $NT\alpha(\tau_0)=o(1)$, $T(T+N^{1/2})\alpha(\tau_*)^{(1+\sigma)/(2+\sigma)}=o(1)$, *and* $N^{1/2}T^{-1}\tau_*^2=o(1)$.

(iv) *Let* $u_i \equiv (u_{i1}, \dots, u_{iT})'$. *Then* (X_i, u_i), $i = 1, \dots, N$, *are mutually independent of each other.*

(v) *For each* $i = 1, \dots, N$, $E(u_{it} \mid \mathcal{F}_{NT,t-1}) = 0$ *almost surely* (*a.s.*).

Assumption 2. *Under* $\mathbb{H}_0(K_0)$, *we have*

(i) $N_k/N \to \tau_k \in (0,1)$ *for each* $k = 1, \dots, K_0$ *as* $N \to \infty$,

(ii) $\sum_{k=1}^{K_0} \|\hat{\alpha}_k - \alpha_k^0\|^2 = O_P((NT)^{-1} + T^{-2})$,

(iii) $P(\bigcup_{k=1}^{K_0} \hat{E}_{kNT}) = o(1)$ *and* $P(\bigcup_{k=1}^{K_0} \hat{F}_{kNT}) = o(1)$.

Assumption 3. *There exist finite nonnegative numbers* c_1 *and* c_2 *such that*

$$\limsup_{(N,T)\to\infty} N \log(NT)/T^2 = c_1 \quad \textit{and}$$

$$\limsup_{(N,T)\to\infty} \log(NT) N^{(3+\sigma)/(4+2\sigma)} T^{-(5+3\sigma)/(4+2\sigma)} = c_2.$$

Assumption 1(i) imposes moment conditions on X_{it} and u_{it}. Assumption 1(ii) requires that Ω_i be positive definite uniformly in i. Assumption 1(iii) requires that each individual time series $\{(X_{it}, u_{it}) : t = 1, 2, \dots\}$ be strong mixing. This condition can be verified if X_{it} does not contain lagged dependent variables regardless of whether one treats the fixed effects μ_is as random or fixed. In the case of dynamic panel data models, Hahn and Kuersteiner (2011) assume that μ_is are nonrandom and uniformly bounded, in which case the strong mixing condition can also be verified. In the case of random fixed effects, they suggest adopting the concept of *conditional strong mixing*, where the mixing coefficient is defined by conditioning on the fixed effects. The dependence of the mixing rate on σ defined in Assumption 1(i) reflects the trade-off between the degree of dependence and the moment bounds of the process $\{(X_{it}, u_{it}), t \geq 1\}$. The last set of conditions in Assumption 1(iii) can easily be met. In particular, if the process is strong mixing with a geometric mixing rate, the conditions on $\alpha(\cdot)$ can be met simply by specifying $\tau_0 = \tau_* = \lfloor C_\tau \log T \rfloor$ for some sufficiently large C_τ, where $\lfloor a \rfloor$ denotes the integer part of a. Assumption 1(iv) rules out cross-sectional dependence among (X_i, u_i) and greatly facilitates our asymptotic analysis. Assumption 1(v) requires that the error term u_{it} be a martingale difference sequence (m.d.s.) with respect to the filter $\mathcal{F}_{NT,t}$, which allows for lagged dependent variables in X_{it}, and conditional heteroskedasticity, skewness, or kurtosis of unknown form in u_{it}.[3]

Assumption 2(i) is typically imposed in the literature on panel data models with latent group structure; see BM, Ando and Bai (2016) and SSP who have rigorous asymptotic analysis for clustering estimators in different contexts. It implies that each group

[3] If the error terms are serially correlated in a static panel, it is well known that by adding lagged dependent and independent variables in the model, one can potentially ameliorate problems caused by such serial correlation. See Su and Chen (2013, p. 1090). For dynamic panel data models (e.g., panel AR(1) model), we cannot allow for serial correlation in the error terms (e.g., AR(1) errors); otherwise, the error terms will be correlated with the lagged dependent variables, causing the endogeneity issue, which we do not address in this paper.

has asymptotically nonnegligible members as $N \to \infty$. Assumption 2(ii) and(iii) can be verified for either the SSP's C-Lasso estimators or the K-means algorithm estimators under suitable conditions. Please note neither BM nor Ando and Bai (2016) study the same model as ours. But it is not difficult to extend their analysis to our model and verify Assumption 2(ii) and (iii). Our model is a special case considered by SSP. We can readily verify Assumption 2(ii) and (iii) under Assumption 1 and the additional conditions (i) $N^{1/2}T^{-1}(\ln T)^9 \to 0$ and $N^2 T^{-(3+2\sigma)} \to c \in [0, \infty)$ as $(N, T) \to \infty$, and (ii) $T\lambda^2/(\ln T)^{6+2\upsilon} \to \infty$ and $\lambda(\ln T)^\upsilon \to 0$ for some $\upsilon > 0$ as $(N, T) \to \infty$. Note that we allow X_{it} to include the lagged dependent variables, in which case, both the SSP's post-Lasso estimators $\hat{\alpha}_k$ or the K-means algorithm estimators have asymptotic bias of order $O(T^{-1})$. In the absence of dynamics in the model, one would expect that $\hat{\alpha}_k - \alpha_k^0 = O_P((NT)^{-1/2})$.

Assumption 3 imposes conditions on the rates at which N and T pass to infinity, and the interaction between (N, T) and σ. Note that we allow N and T to pass to infinity at either identical or suitably restricted different rates. The appearance of the logarithm terms is due to the use of a Bernstein inequality for strong mixing processes. If the mixing process $\{(X_{it}, u_{it}), t \geq 1\}$ has a geometric decay rate, one can take an arbitrarily small σ in Assumption 1(i). In this case, Assumption 3 puts the most stringent restrictions on (N, T) by passing $\sigma \to 0$: $N^{3/5}/T \to 0$ as $(N, T) \to \infty$, ignoring the logarithm term. On the other hand, if $\sigma \geq 1$ in Assumption 1(i), then the second condition in Assumption 3 becomes redundant given the first condition. In the case of conventional panel data models with strictly exogenous regressors only, Pesaran and Yamagata (2008) require that either $\sqrt{N}/T \to 0$ or $\sqrt{N}/T^2 \to 0$ for two of their tests; but for stationary dynamic panel data models, they prove the asymptotic validity of their test only under the condition that $N/T \to \kappa \in [0, \infty)$.

3.2 *Asymptotic null distribution*

Let $h_{i,ts}$ denote the (t, s)th element of $H_i \equiv M_0 X_i (X_i' M_0 X_i)^{-1} X_i' M_0$. Let $X_{it}^\dagger \equiv X_{it} - T^{-1}\sum_{s=1}^T E(X_{is})$ and $\bar{b}_{it} \equiv \Omega_i^{-1/2} X_{it}^\dagger$. Define

$$B_{NT} \equiv N^{-1/2}\sum_{i=1}^N \sum_{t=1}^T u_{it}^2 h_{i,tt} \quad \text{and} \quad V_{NT} \equiv 4T^{-2}N^{-1}\sum_{i=1}^N \sum_{t=2}^T E\left[u_{it}\bar{b}_{it}' \sum_{s=1}^{t-1} \bar{b}_{is} u_{is}\right]^2. \tag{3.2}$$

The following theorem studies the asymptotic null distribution of the infeasible statistic LM_{1NT}.

Theorem 3.1. *Suppose Assumptions 1–3 hold. Then under* $\mathbb{H}_0(K_0)$,

$$J_{1NT}(K_0) \equiv \big(N^{-1/2}\,\mathrm{LM}_{1NT}(K_0) - B_{NT}\big)/\sqrt{V_{NT}} \xrightarrow{D} N(0, 1) \quad \text{as } (N, T) \to \infty.$$

The proof of the above theorem is tedious and is relegated to the Appendix. To implement the test, we need consistent estimates of both B_{NT} and V_{NT}. Let $\hat{b}_{it} = \hat{\Omega}_i^{-1/2}(X_{it} -$

$T^{-1}\sum_{s=1}^{T}X_{is})$ and $\hat{b}_i \equiv (\hat{b}_{i1},\dots,\hat{b}_{iT})'$. Note that $\hat{b}_i = M_0 X_i \hat{\Omega}_i^{-1/2}$. We propose to estimate B_{NT} by

$$\hat{B}_{NT}(K_0) = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2 h_{i,tt} \tag{3.3}$$

and V_{NT} by

$$\hat{V}_{NT}(K_0) = 4T^{-2}N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\hat{u}_{it}\hat{b}_{it}'\sum_{s=1}^{t-1}\hat{b}_{is}\hat{u}_{is}\right]^2. \tag{3.4}$$

Define

$$\hat{J}_{1NT}(K_0) \equiv \left(N^{-1/2}\,\mathrm{LM}_{1NT}(K_0) - \hat{B}_{NT}(K_0)\right)/\sqrt{\hat{V}_{NT}(K_0)}. \tag{3.5}$$

The following theorem establishes the consistency of $\hat{B}_{NT}(K_0)$ and $\hat{V}_{NT}(K_0)$ and the asymptotic distribution of $\hat{J}_{1NT}(K_0)$ under $\mathbb{H}_0(K_0)$.

Theorem 3.2. *Suppose Assumptions 1–3 hold. Then under* $\mathbb{H}_0(K_0)$, $\hat{B}_{NT}(K_0) = B_{NT} + o_P(1)$, $\hat{V}_{NT}(K_0) = V_{NT} + o_P(1)$, *and* $\hat{J}_{1NT}(K_0) \xrightarrow{D} N(0,1)$ *as* $(N,T)\to\infty$.

Theorem 3.2 implies that the test statistic $\hat{J}_{1NT}(K_0)$ is asymptotically pivotal under $\mathbb{H}_0(K_0)$ and we reject $\mathbb{H}_0(K_0)$ for a sufficiently large value of $\hat{J}_{1NT}(K_0)$.

We obtain the distributional results in Theorems 3.1 and 3.2 despite the fact that the individual effects μ_i can only be estimated at the slower rate $T^{-1/2}$ rather than the rate $(NT)^{-1/2}$ or $(NT)^{-1/2}+T^{-1}$ at which the group-specific parameter estimates $\{\hat{\alpha}_k, k=1,\dots,K_0\}$ converge to their true values under $\mathbb{H}_0(K_0)$. The slow convergence rate of these individual effect estimates does not have adverse asymptotic effects on the estimation of the bias term B_{NT}, the variance term V_{NT}, and the asymptotic distribution of $\hat{J}_{1NT}(K_0)$. Nevertheless, they can play an important role in finite samples, which we verify through Monte Carlo simulations.

3.3 *Consistency*

Let $\mathcal{G}_K = \{(G_1,\dots,G_K) : \bigcup_{k=1}^{K} G_k = \{1,\dots,N\}$ and $G_k \cap G_j = \emptyset$ for any $j \neq k\}$. That is, $\mathcal{G}_K$ denotes the class of all possible K-group partitions of $\{1,\dots,N\}$. To study the consistency of our test, we add the following assumption.

Assumption 4. (i) *We have* $N^{-1}\sum_{i=1}^{N}\|\beta_i^0\|^2 = O_P(1)$.

(ii) *We have* $\inf_{(G_1,\dots,G_{K_0})\in\mathcal{G}_{K_0}}\min_{(\alpha_1,\dots,\alpha_{K_0})} N^{-1}\sum_{k=1}^{K_0}\sum_{i\in G_k}\|\beta_i^0-\alpha_k\|^2 \to^P \underline{c}_{K_0} > 0$ *as* $N\to\infty$.

Assumption 4(i) is trivially satisfied if β_i^0s are uniformly bounded or random with finite second moments. Assumption 4(ii) essentially says that one cannot group the N parameter vectors $\{\beta_i^0, 1\le i\le N\}$ into K_0 groups by leaving out an insignificant number of unclassified individuals. It is satisfied for a variety of global alternatives:

(i) The number of groups is $K = K_0 + r$ for some positive integer r such that $N_k/N \to \tau_k \in (0, 1)$ for each $k = 1, \ldots, K_0 + r$.

(ii) There is no grouped pattern among $\{\beta_i^0, 1 \leq i \leq N\}$ such that we have a completely heterogeneous population of individuals.

(iii) The regression model is actually a random coefficient model $\beta_i^0 = \beta^0 + v_i$, where β^0 is a fixed parameter vector, and v_is are independent and identical draws from a continuous distribution with zero mean and finite variance.

(iv) The regression model is a hierarchical random coefficient model $\beta_i^0 = \sum_{k=1}^K (\alpha_k^0 + v_{ki}) \times \mathbf{1}\{i \in G_k^0\}$, where $\alpha_1^0, \ldots, \alpha_K^0$ are defined as before, v_{ki}s (for $k = 1, \ldots, K$) are independent and identical draws from a continuous distribution with zero mean and finite variance, and K may be different from K_0.

The following theorem establishes the consistency of $\hat{J}_{1NT}$.

Theorem 3.3. *Suppose Assumptions 1, 3, and 4 hold. Then under $\mathbb{H}_1(K_0)$ with possible diverging $K_{\max}$ and random coefficients, $P(\hat{J}_{1NT}(K_0) \geq c_{NT}) \to 1$ as $(N, T) \to \infty$ for any nonstochastic sequence $c_{NT} = o(N^{1/2}T)$.*

The above theorem indicates that our test statistic $\hat{J}_{1NT}(K_0)$ is divergent at $N^{1/2}T$ rate under $\mathbb{H}_1(K_0)$ and thus has the power to detect any alternatives such that Assumption 4 is satisfied.

Remark 3.1. Our asymptotic theories here are "pointwise,"as the unknown parameters are treated as fixed. We do not consider the uniformity issue, and so our procedures may suffer from the same problem as the other post-selection inference (see, e.g., Leeb and Pötscher (2005, 2008, 2009) and Schneider and Pötscher (2009)). This seems to be a well known challenge in the literature of model selection or pretest estimation. Although it is a very important question, developing a thorough theory on uniform inference is beyond the scope of this paper.

4. Time fixed effects

In applications, we may also consider the model with both individual and time fixed effects,

$$y_{it} = \beta_i^{0\prime} X_{it} + \mu_i + \gamma_t + u_{it}, \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \tag{4.1}$$

where γ_t is the time fixed effects, the other variables are defined as above, and we assume that β_i^0s have the unknown group structure defined in (2.2). We treat μ_i and γ_t as unknown fixed parameters that are not separately identifiable. Despite this, it is possible to estimate the group-specific parameters α_k^0 and identify each individual's group membership consistently.

4.1 *Estimation under the null and test statistic*

When $\beta_i^0 = \beta^0$ for all $i = 1, \dots, N$, one can follow Hsiao (2014, Chapter 3.6) and sweep out both the individual and time effects from (4.1) via suitable transformation. We do a similar thing in the presence of heterogeneous slope coefficients β_i^0. As before, we first eliminate the individual effects μ_i in (4.1) via the within-group transformation

$$\tilde{y}_{it} = \beta_i' \tilde{X}_{it} + \tilde{\gamma}_t + \tilde{u}_{it}, \tag{4.2}$$

where $\tilde{u}_{it} = u_{it} - \bar{u}_{i\cdot}$, $\tilde{\gamma}_t = \gamma_t - \bar{\gamma}$, and $\bar{\gamma} = T^{-1} \sum_{t=1}^T \gamma_t$. Then we eliminate $\tilde{\gamma}_t$ from the above model to obtain

$$\ddot{y}_{it} = \beta_i' \tilde{X}_{it} - \frac{1}{N} \sum_{j=1}^N \beta_j' \tilde{X}_{jt} + \ddot{u}_{it}, \tag{4.3}$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}$, $\bar{y}_{\cdot t} = \frac{1}{N} \sum_{i=1}^N y_{it}$, $\bar{y} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it}$, and $\ddot{u}_{it}$, $\bar{u}_{\cdot t}$, and $\bar{u}$ are similarly defined. Under $\mathbb{H}_0(K_0)$: $K = K_0$, we can follow SSP and estimate $\boldsymbol{\beta} \equiv (\beta_1^0, \dots, \beta_N^0)$ and $\boldsymbol{\alpha}_{K_0} \equiv (\alpha_1^0, \dots, \alpha_{K_0}^0)$ by minimizing the criterion function

$$Q_{2NT,\lambda}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}_{K_0}) = Q_{2,NT}(\boldsymbol{\beta}) + \frac{\lambda}{N} \sum_{i=1}^N \prod_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \tag{4.4}$$

where

$$Q_{2,NT}(\boldsymbol{\beta}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(\ddot{y}_{it} - \beta_i' \tilde{X}_{it} + \frac{1}{N} \sum_{j=1}^N \beta_j' \tilde{X}_{jt} \right)^2.$$

Let $\tilde{\boldsymbol{\beta}} \equiv (\tilde{\beta}_1, \dots, \tilde{\beta}_N)$ and $\tilde{\boldsymbol{\alpha}}_{K_0} \equiv (\tilde{\alpha}_1, \dots, \tilde{\alpha}_{K_0})$ denote the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_{K_0}$, respectively. If necessary, we can estimate $\tilde{\gamma}_t$ by $\tilde{\tilde{\gamma}}_t = \frac{1}{N} \sum_{i=1}^N (\tilde{y}_{it} - \tilde{\beta}_i' \tilde{X}_{it})$ for $t = 1, \dots, T$. Given $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\alpha}}_{K_0}$, we classify the N individuals into K_0 groups as in Section 2. As before, we use $\hat{G}_1, \dots, \hat{G}_{K_0}$ to denote the K_0 estimated groups and $\hat{N}_k$ to denote the cardinality of $\hat{G}_k$.

The post-Lasso estimates can be obtained in two ways. One is to pool all the observations within each estimated group and estimate the group-specific parameters for each group separately after we demean over time and across individuals within the group. In this way, one can easily work out the standard error for each group-specific estimate as usual. Alternatively, we consider estimation based on (4.3) by using the estimated group membership. Assuming all individuals are classified into one of the K_0 groups, we obtain the post-Lasso estimates $\{\hat{\alpha}_k\}$ of $\{\alpha_k^0\}$ as the solution to the minimization problem

$$\min_{\{\alpha_k\}} \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k} \sum_{t=1}^T \left(\ddot{y}_{it} - \alpha_k' \tilde{X}_{it} + \frac{1}{N} \sum_{l=1}^{K_0} \sum_{j \in \hat{G}_k} \alpha_j' \tilde{X}_{jt} \right)^2. \tag{4.5}$$

Let $\hat{\beta}_i = \hat{\alpha}_k$ for $i \in \hat{G}_k$ and $k = 1, \dots, K_0$. Let $\hat{\boldsymbol{\alpha}}_{K_0} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{K_0})$.

Given the above post-Lasso estimates, we define the restricted residual as

$$\widehat{\widehat{u}}_{it} = \ddot{y}_{it} - \hat{\beta}_i' \tilde{X}_{it} + \frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_j' \tilde{X}_{jt}. \tag{4.6}$$

Following the analysis in Section 2.2, we propose to test $\mathbb{H}_0(K_0)$ by using the LM-type test statistic

$$\mathrm{LM}_{2NT}(K_0) = \sum_{i=1}^{N} \widehat{\widehat{u}}_i' M_0 X_i \big(X_i' M_0 X_i\big)^{-1} X_i' M_0 \widehat{\widehat{u}}_i,$$

where $\widehat{\widehat{u}}_i = (\widehat{\widehat{u}}_{i1}, \ldots, \widehat{\widehat{u}}_{iT})'$. We will show that after being appropriately centered, $N^{-1/2} \times \mathrm{LM}_{2NT}(K_0)$ follows the same asymptotic distribution as $N^{-1/2}\,\mathrm{LM}_{1NT}(K_0)$ under the null.

4.2 *Asymptotic properties of the C-Lasso estimates*

Since SSP do not allow the fixed time effects in their model, their asymptotic theory does not apply to our C-Lasso estimates of the parameters and group identity. Fortunately, we can study the preliminary convergence rates of various estimates under Assumptions 1, 2(i), and 3.

Theorem 4.1. *Suppose that Assumptions* 1, 2(i), *and* 3 *hold. Then under* $\mathbb{H}_0(K_0)$,

(i) $\frac{1}{N}\sum_{i=1}^{N} \|\tilde{\beta}_i - \beta_i^0\|^2 = O_P(T^{-1})$,

(ii) $(\tilde{\alpha}_{(1)}, \ldots, \tilde{\alpha}_{(K_0)}) - (\alpha_1^0, \ldots, \alpha_{K_0}^0) = O_P(T^{-1/2})$,

(iii) $\max_{1\le i\le N} \|\tilde{\beta}_i - \beta_i^0\| = O_P(a_{NT} + \lambda)$,

where $a_{NT} = \max\{(NT)^{1/(4+2\sigma)} \log(NT)/T, (\log(NT)/T)^{1/2}\}$ *and* $(\tilde{\alpha}_{(1)}, \ldots, \tilde{\alpha}_{(K_0)})$ *is some suitable permutation of* $(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_{K_0})$.

Theorem 4.1(i)–(iii) establishes the mean square convergence rate of $\{\tilde{\beta}_i\}$, the convergence rate of the group-specific estimates $\{\tilde{\alpha}_k\}$, and the uniform convergence rate of $\tilde{\beta}_i$s, respectively. The findings are similar to those in SSP. In particular, the estimates of the group-specific parameters do not depend on λ. Note that $a_{NT} = O((\log(T)/T)^{1/2})$ under the additional restriction that $\limsup_{(N,T)\to\infty} NT^{-(1+\sigma)} = c \in [0, \infty)$. For notational simplicity, hereafter we simply write $\tilde{\alpha}_k$ for $\tilde{\alpha}_{(k)}$ as the consistent estimator of α_k^0s.

Define the two types of classification errors $\hat{E}_{kNT}$ and $\hat{F}_{kNT}$ as before. To study the consistency of our classification method, we add the following assumption.

Assumption 5. (i) *We have* $N^{1/2}T^{-1}(\ln T)^9 \to 0$ *and* $\limsup_{(N,T)\to\infty} NT^{-(1+\sigma)} = c \in [0, \infty)$.

(ii) *We have* $T\lambda^2/(\ln T)^{6+2\upsilon} \to \infty$ *and* $\lambda(\ln T)^{\upsilon} \to 0$ *for some* $\upsilon > 0$ *as* $(N, T) \to \infty$.

Assumption 5(i) and (ii) strengthens the conditions on N, T, and λ. Note that Assumption 5(ii) implies that it suffices to choose $\lambda \propto T^{-a}$ for some $a \in (1/4, 1/2)$.

The following theorem establishes the uniform consistency for our classification method.

Theorem 4.2. *Suppose that Assumptions 1, 2(i), 3, and 5 hold. Then under* $\mathbb{H}_0(K_0)$,

(i) $P(\bigcup_{k=1}^{K_0} \hat{E}_{kNT}) \le \sum_{k=1}^{K_0} P(\hat{E}_{kNT}) \to 0$ *as* $(N, T) \to \infty$,

(ii) $P(\bigcup_{k=1}^{K_0} \hat{F}_{kNT}) \le \sum_{k=1}^{K_0} P(\hat{F}_{kNT}) \to 0$ *as* $(N, T) \to \infty$.

Given the results in Theorems 4.1 and 4.2, we will show that the post-Lasso estimators $\{\hat{\alpha}_k\}$ are asymptotically oracle efficient in the sense that they are as efficient as the infeasible estimators $\{\bar{\alpha}_k\}$ under some regularity conditions:

$$\{\bar{\alpha}_k\} \equiv \arg\min_{\{\alpha_k\}} \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i\in G_k^0} \sum_{t=1}^{T} \left(\ddot{y}_{it} - \alpha_k' \tilde{X}_{it} + \frac{1}{N} \sum_{l=1}^{K_0} \sum_{j \in G_k^0} \alpha_j' \tilde{X}_{jt} \right)^2. \tag{4.7}$$

Let $\bar{\boldsymbol{\alpha}}_{K_0} = (\bar{\alpha}_1, \ldots, \bar{\alpha}_{K_0})$. Let $\mathbb{Q}_{kNT} = \frac{1}{N_k T} \sum_{i\in G_k^0} \tilde{\mathbb{X}}_i' \tilde{X}_i$, $\mathbb{Q}_{k,l} = \frac{1}{N N_k T} \sum_{i \in G_k^0} \sum_{j \in G_l^0} \tilde{\mathbb{X}}_i' \tilde{X}_j$, and $\mathbb{V}_{kNT} = \frac{1}{N_k T} \sum_{i\in G_k^0} \tilde{\mathbb{X}}_i' \ddot{y}_i$ for $k, l = 1, \ldots, K_0$, where $\tilde{\mathbb{X}}_i = \tilde{X}_i - N^{-1} \sum_{j\in G_k^0} \tilde{X}_j$ and $\ddot{y}_i = (\ddot{y}_{i1}, \ldots, \ddot{y}_{iT})'$. Define

$$\mathbb{Q}_{NT} = \begin{pmatrix} \mathbb{Q}_{1NT} - \mathbb{Q}_{1,1} & -\mathbb{Q}_{1,2} & \cdots & -\mathbb{Q}_{1,K_0} \\ -\mathbb{Q}_{2,1} & \mathbb{Q}_{2NT} - \mathbb{Q}_{2,2} & \cdots & -\mathbb{Q}_{2,K_0} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbb{Q}_{K_0,1} & -\mathbb{Q}_{K_0,2} & \cdots & \mathbb{Q}_{K_0NT} - \mathbb{Q}_{K_0,K_0} \end{pmatrix} \quad \text{and}$$
$$\mathbb{V}_{NT} = \begin{pmatrix} \mathbb{V}_{1NT} \\ \mathbb{V}_{2NT} \\ \vdots \\ \mathbb{V}_{K_0NT} \end{pmatrix}. \tag{4.8}$$

Let $\hat{\mathbb{Q}}_{NT}$ and $\hat{\mathbb{V}}_{NT}$ be analogously defined as $\mathbb{Q}_{NT}$ and $\mathbb{V}_{NT}$ with G_k^0s and $N_k s$ replaced by $\hat{G}_k$s and $\hat{N}_k$s. We can verify that

$$\operatorname{vec}(\bar{\boldsymbol{\alpha}}_{K_0}) = \mathbb{Q}_{NT}^{-1} \mathbb{V}_{NT} \quad \text{and} \quad \operatorname{vec}(\hat{\boldsymbol{\alpha}}_{K_0}) = \hat{\mathbb{Q}}_{NT}^{-1} \hat{\mathbb{V}}_{NT}.$$

Let $\mathbb{B}_{kNT} \equiv \frac{1}{N_k T^2} \sum_{i\in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} X_{is} u_{it}$, $\mathbb{U}_{kNT} \equiv \frac{1}{N_k T} \sum_{i\in G_k^0} \sum_{t=1}^{T} X_{it}^{\perp} u_{it}$, and $\Omega_{kNT} = \frac{N}{N_k^2 T} \sum_{i\in G_k^0} \sum_{t=1}^{T} E(X_{it}^{\perp} X_{it}^{\perp\prime} u_{it}^2)$ for $k = 1, \ldots, K_0$, where $X_{it}^{\perp} = X_{it} - E(\bar{X}_{\cdot t}^{(k)} - \bar{X}^{(k)})$, $\bar{X}_{\cdot t}^{(k)} = \frac{1}{N} \sum_{i\in G_k^0} X_{it}$, and $\bar{X}^{(k)} = \frac{1}{T} \sum_{t=1}^{T} \bar{X}_{\cdot t}^{(k)}$.[4] Let $\mathbb{B}_{NT} = (\mathbb{B}_{1NT}', \ldots, \mathbb{B}_{K_0NT}')'$, $\mathbb{U}_{NT} = (\mathbb{U}_{1NT}', \ldots, \mathbb{U}_{K_0NT}')'$, and $\Omega_{NT} = \operatorname{diag}(\Omega_{1NT}, \ldots, \Omega_{K_0NT})$.

We add the following assumption.

[4] If $\{X_{it}, t \ge 1\}$ is mean-stationary, then $X_{it}^{\perp} = X_{it}$.

ASSUMPTION 6. (i) *We have* $\mathbb{Q}_{NT} \to^P \mathbb{Q}_0 > 0$ *as* $(N, T) \to \infty$.

(ii) *We have* $\sqrt{NT}\mathbb{U}_{NT} \to^D N(0, \Omega_0)$ *as* $(N, T) \to \infty$ *where* $\Omega_0 = \lim_{(N,T)\to\infty} \Omega_{NT}$.

Assumption 6 imposes conditions to ensure the asymptotic normality of the post-Lasso estimators of the group-specific parameters. When $K_0 = 1$, $\mathbb{Q}_{NT}$ reduces to $\mathbb{Q}_{1NT} - \mathbb{Q}_{1,1}$, which is required to be asymptotically nonsingular so that the usual fixed effects estimator in a two-way error component model is asymptotically normal.

THEOREM 4.3. *Suppose Assumptions* 1, 2(i), 3, 5, *and* 6 *hold. Then under* $\mathbb{H}_0(K_0)$,

(i) $\hat{\alpha}_k - \alpha_k^0 = \bar{\alpha}_k - \alpha_k^0 + o_P((NT)^{-1/2}) = O_P((NT)^{-1/2} + T^{-1})$ *for* $k = 1, \ldots, K_0$,

(ii) $\sqrt{NT}[\text{vec}(\hat{\boldsymbol{\alpha}}_{K_0} - \boldsymbol{\alpha}_{K_0}^0) + \mathbb{Q}_{NT}^{-1}\mathbb{B}_{NT}] \to^D N(0, \mathbb{Q}_0^{-1}\Omega_0\mathbb{Q}_0^{-1})$.

Theorem 4.3(i) indicates that $\hat{\alpha}_k$ is asymptotically equivalent to the infeasible estimator $\bar{\alpha}_k$. It also reports the convergence rate of $\hat{\alpha}_k$ and implies that $\frac{1}{N}\sum_{i=1}^N \|\hat{\beta}_i - \beta_i^0\|^2 = O_P((NT)^{-1} + T^{-2})$. Theorem 4.3(ii) reports the asymptotic distribution of $\hat{\boldsymbol{\alpha}}_{K_0}$. By the Davydov inequality, we can readily show that $\mathbb{Q}_{NT}^{-1}\mathbb{B}_{NT}$ is $O_P(T^{-1})$, which suggests that the order of the asymptotic bias of $\text{vec}(\hat{\boldsymbol{\alpha}}_{K_0})$ is given by $O(T^{-1})$. Note that this bias term has the exact form as the bias term in the linear panel data model with individual fixed effects alone considered by SSP. As in SSP, the bias is asymptotically vanishing in the case where X_{it} only contains strictly exogenous regressors or $N = o(T)$.

For inference, one needs to estimate $\mathbb{Q}_0$, Ω_0, and $\mathbb{Q}_{NT}^{-1}\mathbb{B}_{NT}$. When necessary, one can follow SSP to estimate $\mathbb{Q}_{NT}^{-1}\mathbb{B}_{NT}$. A consistent estimate of $\mathbb{Q}_0$ is given by $\hat{\mathbb{Q}}_{NT}$. Similarly, we can consistently estimate Ω_0 by a sample analogue of Ω_{NT}. For brevity, we do not give the details.

4.3 *Asymptotic null distribution of* $\text{LM}_{2NT}(K_0)$

Despite the presence of time fixed effects, we can show that $\text{LM}_{2NT}(K_0)$ shares the same asymptotic null distribution as $\text{LM}_{1NT}(K_0)$. This is summarized in the following theorem.

THEOREM 4.4. *Suppose Assumptions* 1, 2(i), 3, 5, *and* 6 *hold. Then under* $\mathbb{H}_0(K_0)$,

$$J_{2NT}(K_0) \equiv \big(N^{-1/2}\,\text{LM}_{2NT}(K_0) - B_{NT}\big)/\sqrt{V_{NT}} \xrightarrow{D} N(0,1) \quad \textit{as } (N, T) \to \infty.$$

That is, $N^{-1/2}\,\text{LM}_{2NT}(K_0)$ shares the same asymptotic bias (B_{NT}) and asymptotic variance (V_{NT}) as $N^{-1/2}\,\text{LM}_{1NT}(K_0)$.

As before, we can consistently estimate both B_{NT} and V_{NT} by $\hat{B}_{NT}(K_0)$ and $\hat{V}_{NT}(K_0)$. The major difference is that we now need to use the residuals $\widehat{\hat{u}}_{it}$ to replace $\hat{u}_{it}$ throughout. Given $\hat{B}_{NT}(K_0)$ and $\hat{V}_{NT}(K_0)$, we can obtain a feasible version of $J_{2NT}(K_0)$:

$$\hat{J}_{2NT}(K_0) \equiv \big(N^{-1/2}\,\text{LM}_{2NT}(K_0) - \hat{B}_{NT}(K_0)\big)/\sqrt{\hat{V}_{NT}(K_0)}.$$

Following the proofs of Theorems 3.2 and 3.3, we can readily show that $\hat{J}_{2NT}(K_0)$ is asymptotically $N(0, 1)$ under $\mathbb{H}_0(K_0)$ and is divergent at $N^{1/2}T$ rate under under $\mathbb{H}_1(K_0)$.

That is, results analogous to those in Theorems 3.2 and 3.3 also hold. For brevity we omit the details.

5. Monte Carlo simulations

In this section, we conduct Monte Carlo simulations to examine the finite-sample performance of our proposed testing method.

5.1 *Data generating processes and implementation*

We consider five generating processes (DGPs). The first four DGPs only have individual fixed effects:

DGP 1: $y_{it} = \beta^0_{1i} X_{1it} + \beta^0_{2i} X_{2it} + \mu_i + u_{it}$.

DGPs 2–4: $y_{it} = \beta^0_{1i} X_{1it} + \beta^0_{2i} y_{i,t-1} + \mu_i + u_{it}$, where $X_{jit} = \xi_{jit} + \mu_i$, $j = 1, 2$, and μ_i, ξ_{1it}, ξ_{2it}, and u_{it} are independent and identically distributed (IID) $N(0, 1)$ variables and are mutually independent of each other.

DGP 5 has both individual and time fixed effects:

DGP 5: $y_{it} = \beta^0_{1i} X_{1it} + \beta^0_{2i} y_{i,t-1} + 0.5\mu_i + 0.5\gamma_t + u_{it}$, where $X_{1it} = \xi_{1it} + \mu_i + \gamma_t$, and μ_i, γ_t, ξ_{1it}, and u_{it} are mutually independent IID $N(0, 1)$ variables.

DGP 1 is a static panel structure, while DGPs 2–5 are dynamic panel structures. In DGPs 1, 2 and 5, $(\beta^0_{1i}, \beta^0_{2i})$ has a group structure:

$$\left(\beta^0_{1i}, \beta^0_{2i}\right) = \begin{cases} (0.5, -0.5) & \text{with probability } 0.3, \\ (-0.5, 0.5) & \text{with probability } 0.3, \\ (0, 0) & \text{with probability } 0.4. \end{cases}$$

Therefore, in DGPs 1, 2, and 5, the true number of groups is three. In DGP 3, we consider a completely heterogeneous (random coefficient) panel structure where β^0_{1i} and β^0_{2i} follow $N(0.5, 1)$ and $\mathrm{U}(-0.5, 0.5)$, respectively. In principle, the true number of groups is the cross-sectional dimension N in this case. In DGP 4, $(\beta^0_{1i}, \beta^0_{2i})$ is similar to that in DGPs 1, 2, and 5 except that it has some additional small disturbance. Specifically,

$$\left(\beta^0_{1i}, \beta^0_{2i}\right) = \begin{cases} (0.5 + 0.1\nu_{1i}, -0.5 + 0.1\nu_{2i}) & \text{with probability } 0.3, \\ (-0.5 + 0.1\nu_{1i}, 0.5 + 0.1\nu_{2i}) & \text{with probability } 0.3, \\ (0.1\nu_{1i}, 0.1\nu_{2i}) & \text{with probability } 0.4, \end{cases}$$

where ν_{1i} and ν_{2i} are each IID $N(0, 1)$, mutually independent, and independent of μ_i, ξ_{it}, and u_{it}. DGP 4 can be thought of as a small deviation from a group structure.

For each DGP, we first test the null hypotheses $\mathbb{H}_0(1)$, $\mathbb{H}_0(2)$, and $\mathbb{H}_0(3)$ to examine the level and power of our test.

We then use our tests to determine the number of groups as described in Section 2.1. We set $K_{\max} = 8$ and let the nominal size decrease with the time series dimension T to ensure that the type I error decreases with T. Specifically, we let the nominal size be

$1/T$, which equals 0.10 and 0.025 for $T = 10$ and 40, respectively.[5] If all eight hypotheses, $\mathbb{H}_0(1), \ldots, \mathbb{H}_0(8)$, are rejected, then we stop and conclude that the number of groups is greater than eight and we can adopt the random coefficient model.

For the combination of N and T, we consider the typical case in empirical applications that T is smaller than or comparable to N and let $(N, T) = (40, 10)$, $(80, 10)$, and $(40, 40)$. The number of replications in the simulations is 1000.

One important step in implementing our testing procedure is to choose the tuning parameter λ. Following the theory in SSP, we let $\lambda = c \cdot s_Y^2 \cdot T^{-1/3}$, where s_Y is the sample standard deviation of Y_{it} and c is some constant. We use three different values of c (0.25, 0.5, and 0.75) to examine the sensitivity of our results to c (thus λ).

As a comparison, we also consider the information criterion (IC) proposed in SSP to determine the number of groups. Specifically, we choose K to minimize the IC,

$$\text{IC}(K, \lambda) = \ln\left(\hat{\sigma}^2_{(K,\lambda)}\right) + \rho_{1NT}\, pK, \quad K = 1, \ldots, K_{\max}, \tag{5.1}$$

where $\hat{\sigma}^2_{(K,\lambda)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2$, and $\hat{u}_{it}'$s are the residuals under the specification of K groups with the tuning parameter λ. The $\hat{u}_{it}'$s are defined in (2.6) or (4.6) in the presence of both individual and time fixed effects. The notation ρ_{1NT} is a tuning parameter and we set $\rho_{1NT} = \frac{2}{3}(NT)^{-1/2}$ as in SSP. Note that SSP's IC can also be used to choose the tuning parameter λ and K jointly by minimizing $\text{IC}(K, \lambda)$ over both λ and K.

5.2 *Simulation results*

Table 1 shows the level and power behavior of our test statistics for testing the three null hypotheses: $\mathbb{H}_0(1)$, $\mathbb{H}_0(2)$, and $\mathbb{H}_0(3)$. We choose three conventional nominal levels: 0.01, 0.05, and 0.10. For DGPs 1, 2, and 5, the true number of groups is three. For $\mathbb{H}_0(1)$, the rejection frequencies are all 1 for all combinations of N and T and all three DGPs at all three nominal levels. For $\mathbb{H}_0(2)$, the power of the test increases rapidly with both

TABLE 1. Empirical rejection frequency.

				$c = 0.25$			$c = 0.5$			$c = 0.75$		
		N	T	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
DGP 1	$K = 1$	40	10	1	1	1	1	1	1	1	1	1
	(alternative)	80	10	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
	$K = 2$	40	10	0.13	0.31	0.43	0.12	0.30	0.41	0.12	0.29	0.41
	(alternative)	80	10	0.28	0.53	0.67	0.28	0.52	0.66	0.26	0.52	0.65
		40	40	1	1	1	1	1	1	1	1	1
	$K = 3$	40	10	0.000	0.01	0.04	0.01	0.03	0.06	0.01	0.06	0.10
	(null)	80	10	0.001	0.01	0.04	0.003	0.03	0.07	0.02	0.06	0.12
		40	40	0.003	0.02	0.04	0.01	0.03	0.07	0.02	0.06	0.12

(*Continues*)

[5] We also try fixing the nominal level at 0.05. The results are similar and available upon request.

Table 1. *Continued.*

		N	T	$c = 0.25$			$c = 0.5$			$c = 0.75$		
				0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
DGP 2	$K = 1$	40	10	1	1	1	1	1	1	1	1	1
	(alternative)	80	10	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
	$K = 2$	40	10	0.06	0.21	0.32	0.06	0.18	0.30	0.05	0.17	0.29
	(alternative)	80	10	0.16	0.38	0.54	0.14	0.36	0.51	0.14	0.35	0.50
		40	40	1	1	1	1	1	1	1	1	1
	$K = 3$	40	10	0.001	0.01	0.01	0.001	0.01	0.02	0.001	0.02	0.04
	(null)	80	10	0.000	0.003	0.01	0.002	0.01	0.02	0.001	0.02	0.05
		40	40	0.001	0.01	0.03	0.01	0.02	0.04	0.01	0.05	0.09
DGP 3	$K = 1$	40	10	1	1	1	1	1	1	1	1	1
	(alternative)	80	10	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
	$K = 2$	40	10	1	1	1	1	1	1	1	1	1
	(alternative)	80	10	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
	$K = 3$	40	10	0.98	1	1	0.98	1	1	0.99	1	1
	(alternative)	80	10	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
DGP 4	$K = 1$	40	10	1	1	1	1	1	1	1	1	1
	(alternative)	80	10	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
	$K = 2$	40	10	0.12	0.30	0.44	0.12	0.28	0.42	0.10	0.27	0.41
	(alternative)	80	10	0.34	0.60	0.72	0.32	0.57	0.70	0.31	0.53	0.69
		40	40	1	1	1	1	1	1	1	1	1
	$K = 3$	40	10	0.001	0.02	0.04	0.004	0.02	0.05	0.004	0.03	0.08
	(alternative)	80	10	0.01	0.05	0.08	0.02	0.06	0.10	0.02	0.08	0.15
		40	40	0.26	0.49	0.62	0.32	0.54	0.67	0.38	0.62	0.74
DGP 5	$K = 1$	40	10	1	1	1	1	1	1	1	1	1
	(alternative)	80	10	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
	$K = 2$	40	10	0.38	0.61	0.75	0.36	0.60	0.74	0.37	0.61	0.74
	(alternative)	80	10	0.72	0.89	0.94	0.71	0.89	0.94	0.71	0.89	0.94
		40	40	1	1	1	1	1	1	1	1	1
	$K = 3$	40	10	0.01	0.04	0.06	0.01	0.03	0.06	0.01	0.04	0.07
	(null)	80	10	0.01	0.04	0.07	0.01	0.05	0.09	0.03	0.07	0.13
		40	40	0.002	0.02	0.05	0.003	0.02	0.06	0.01	0.03	0.07

Note: Numbers in the main entries are the rejection frequencies under the null or the alternative for three nominal levels: 0.01, 0.05, and 0.10. We consider three values of the constant c in the tuning parameter λ: 0.25, 0.5, and 0.75.

N and T. For example, when $N = 40$ and $T = 40$, the rejection frequencies are all 1 at all three nominal levels. For $\mathbb{H}_0(3)$, we examine the level of our test and find that the rejection frequencies are fairly close to the nominal levels.

For the heterogeneous DGP 3, our test rejects all three hypotheses with the frequencies being 1 or nearly 1 at all three nominal levels. This reflects the power of our test against global alternatives. For DGP 4, which represents a small deviation from the group structure, our test shows reasonable power for large T, though it rejects $K = 3$ with a small frequency when T is small, as expected. Also note that all the testing results are quite robust to the values of c (thus λ).

Table 2A shows how our method performs in terms of determining the number of groups. Specifically, the numbers in the main entries are the proportions of the replications in which the number of groups determined by our method is equal to, less than (<5 in DGP 3), or greater than (>5 in DGPs 1, 2, 4, and 5; >8 in DGP 3) a number. For DGPs 1, 2, and 5, our method determines the correct number of groups (three) with a large probability. For example, when $N = 40$ and $T = 40$, for DGP 5, the number of groups determined by our testing procedure equals the true number (three) with a probability of 0.99. For DGP 3, where the true number of groups is N, our method determines a large number of groups (greater than 8) with a probability of 1 when $N = 40$ and $T = 40$. DGP 4 represents a small deviation from a three-group structure. When the sample size is small ($N = 40$ and $T = 10$), with a high probability, the number of groups determined by our method is two or three. When the sample size increases to $N = 40$ and $T = 40$, our method determines four groups with a probability larger than 0.25. These results are reasonable. Intuitively, if N and T are small, the data can only provide limited information on the underlying DGP, and it is reasonable to apply a small group structure to serve as a good approximation to the true model. As N and T become large, more information on the underlying DGP is revealed, and it is sensible to adopt a larger number of groups to approximate the true model more accurately.

Table 2B shows the performance of SSP's information criterion (IC). Comparing Tables 2A and 2B, we find that for DGPs 1, 2, and 5, the performances of our method and SSP's IC are comparable. However, for DGPs 3 and 4, our method dominates SSP's. Specifically, for DGP 3, where the number of groups is N, SSP tend to determine a small number of groups (less than six even when $N = 40$ and $T = 40$), while our method determines that the number of groups is greater than eight for most cases. For DGP 4, in large

TABLE 2A. Frequency of the number of groups determined by our method.

		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 1	$c=0.25$	40	10	0	0.57	0.40	0.03	0	0
		80	10	0	0.34	0.63	0.03	0	0
		40	40	0	0	0.99	0.01	0	0
	$c=0.5$	40	10	0	0.59	0.35	0.05	0.01	0
		80	10	0	0.34	0.60	0.05	0.01	0
		40	40	0	0	0.98	0.02	0	0
	$c=0.75$	40	10	0	0.59	0.31	0.07	0.02	0.01
		80	10	0	0.35	0.53	0.08	0.04	0.01
		40	40	0	0	0.96	0.03	0.01	0

(*Continues*)

Table 2A. *Continued.*

		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 2	$c=0.25$	40	10	0	0.68	0.30	0.01	0	0
		80	10	0	0.46	0.52	0.01	0	0
		40	40	0	0	1	0	0	0
	$c=0.5$	40	10	0	0.71	0.28	0.02	0	0
		80	10	0	0.49	0.49	0.01	0.01	0
		40	40	0	0	0.99	0.01	0	0
	$c=0.75$	40	10	0	0.72	0.25	0.03	0.01	0
		80	10	0	0.51	0.45	0.04	0.01	0
		40	40	0	0	0.98	0.02	0	0
		N	T	$K<5$	$K=5$	$K=6$	$K=7$	$K=8$	$K>8$
DGP 3	$c=0.25$	40	10	0.02	0.03	0.04	0.05	0.01	0.86
		80	10	0	0	0	0	0	1
		40	40	0	0	0	0	0	1
	$c=0.5$	40	10	0.01	0.01	0.01	0.02	0.01	0.93
		80	10	0	0	0	0	0	1
		40	40	0	0	0	0	0	1
	$c=0.75$	40	10	0	0	0.01	0.01	0.01	0.97
		80	10	0	0	0	0	0	1
		40	40	0	0	0	0	0	1
		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 4	$c=0.25$	40	10	0	0.56	0.39	0.04	0.01	0
		80	10	0	0.28	0.64	0.07	0.01	0
		40	40	0	0	0.62	0.26	0.11	0.01
	$c=0.5$	40	10	0	0.58	0.37	0.04	0.01	0
		80	10	0	0.30	0.60	0.09	0.02	0
		40	40	0	0	0.58	0.28	0.13	0.01
	$c=0.75$	40	10	0	0.59	0.33	0.06	0.02	0
		80	10	0	0.31	0.54	0.11	0.03	0.01
		40	40	0	0	0.49	0.34	0.16	0.01
		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 5	$c=0.25$	40	10	0	0.25	0.69	0.05	0	0
		80	10	0	0.06	0.87	0.06	0.01	0
		40	40	0	0	0.99	0.01	0	0
	$c=0.5$	40	10	0	0.26	0.68	0.05	0	0
		80	10	0	0.06	0.85	0.08	0	0
		40	40	0	0	0.99	0.01	0	0
	$c=0.75$	40	10	0	0.27	0.67	0.06	0.01	0
		80	10	0	0.06	0.81	0.12	0.01	0
		40	40	0	0	0.99	0.01	0	0

Note: Numbers in the main entries are the proportions of the replications in which the determined number of groups is less than, equal to, or greater than a number. We consider three values of the constant c in the tuning parameter λ: 0.25, 0.5, and 0.75.

TABLE 2B. Frequency of the number of groups determined by the information criterion.

		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 1	$c=0.25$	40	10	0	0.49	0.50	0.01	0	0
		80	10	0	0.13	0.81	0.06	0.01	0
		40	40	0	0	1	0	0	0
	$c=0.5$	40	10	0	0.67	0.33	0.01	0	0
		80	10	0	0.28	0.68	0.04	0	0
		40	40	0	0	1	0	0	0
	$c=0.75$	40	10	0	0.80	0.20	0	0	0
		80	10	0	0.46	0.52	0.02	0	0
		40	40	0	0	1	0	0	0
		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 2	$c=0.25$	40	10	0	0.55	0.44	0.01	0	0
		80	10	0	0.17	0.79	0.04	0	0
		40	40	0	0	1	0	0	0
	$c=0.5$	40	10	0	0.70	0.30	0	0	0
		80	10	0	0.31	0.66	0.02	0.01	0
		40	40	0	0	1	0	0	0
	$c=0.75$	40	10	0	0.81	0.19	0	0	0
		80	10	0	0.51	0.46	0.03	0	0
		40	40	0	0	1	0	0	0
		N	T	$K<5$	$K=5$	$K=6$	$K=7$	$K=8$	$K>8$
DGP 3	$c=0.25$	40	10	0.96	0.03	0.01	0	0	0
		80	10	0.95	0.05	0	0	0	0
		40	40	0.67	0.21	0.10	0.02	0	0
	$c=0.5$	40	10	0.99	0.01	0	0	0	0
		80	10	0.98	0.02	0	0	0	0
		40	40	0.81	0.13	0.05	0.01	0	0
	$c=0.75$	40	10	0.99	0.01	0	0	0	0
		80	10	0.99	0.01	0	0.00	0	0
		40	40	0.85	0.11	0.02	0.01	0	0
		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 4	$c=0.25$	40	10	0	0.51	0.46	0.03	0	0
		80	10	0	0.16	0.74	0.10	0.01	0
		40	40	0	0	1	0	0	0
	$c=0.5$	40	10	0	0.65	0.34	0.02	0	0
		80	10	0	0.30	0.65	0.05	0.01	0
		40	40	0	0	1	0	0	0
	$c=0.75$	40	10	0	0.77	0.22	0.01	0	0
		80	10	0	0.50	0.45	0.04	0.01	0
		40	40	0	0	0.99	0.01	0	0

(*Continues*)

TABLE 2B. *Continued.*

		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 5	$c=0.25$	40	10	0	0.16	0.79	0.06	0	0
		80	10	0	0.01	0.89	0.10	0.01	0
		40	40	0	0	1	0	0	0
	$c=0.5$	40	10	0	0.18	0.75	0.07	0	0
		80	10	0	0.02	0.83	0.16	0	0
		40	40	0	0	1	0	0	0
	$c=0.75$	40	10	0	0.22	0.70	0.08	0	0
		80	10	0	0.03	0.70	0.27	0	0
		40	40	0	0	1	0	0	0

Note: Numbers in the main entries are the proportions of the replications in which the determined number of groups is less than, equal to, or greater than a number. We consider three values of the constant c in the tuning parameter λ: 0.25, 0.5, and 0.75.

samples ($N=40$ and $T=40$), SSP determine three groups with a high probability, while our method determines a relatively large number of groups.

To examine the performance of the C-Lasso estimator following our method to determine the number of groups, we compare the following four estimators: (i) the common coefficient estimator, (ii) the random coefficient estimator, (iii) the post C-Lasso estimator with the number of groups determined by our method, and (iv) the infeasible estimator where the true number of groups and group classification are known. We report the mean squared errors (MSE), calculated as the average values of $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i-\beta_i^0\|^2$ over 1000 replications, where $\hat{\beta}_i$ denotes the estimator. Table 3 presents the results for DGP 1.[6] It is clear that our estimator dominates the common coefficient estimator and the random coefficient estimator. However it performs worse than the infeasible estimator, especially when the sample size is small. This reflects the effect of classification errors in finite samples.

6. Empirical application: Income and democracy

The relationship between income and democracy has attracted much attention in empirical research; see, for example, Lipset (1959), Barro (1999), AJRY, and BM. To the best

TABLE 3. Comparison of the MSE of various estimators: DGP 1.

				Post C-Lasso			
N	T	Common Coefficient	Random Coefficient	$c=0.25$	$c=0.5$	$c=0.75$	Infeasible
40	10	0.31	0.33	0.18	0.18	0.18	0.02
80	10	0.30	0.33	0.16	0.17	0.17	0.01
40	40	0.30	0.06	0.02	0.02	0.03	0.004

Note: Numbers in the main entries are the MSE of various estimators. For the post C-Lasso estimators, we consider three values of the constant c in the tuning parameter λ: 0.25, 0.5, and 0.75.

[6]The results for other DGPs are available upon request.

of our knowledge, none of the existing studies allows for heterogeneity in the slope coefficients in their model specifications. As discussed in AJRY, "societies may embark on divergent political-economic development paths." Thus, ignoring the heterogeneity in the slope coefficients may result in model misspecification and invalidate subsequent inferences. Hence, it is important to know whether the data support the assumption of homogeneous slope coefficients. If not, then we need to determine the number of heterogeneous groups and classify the countries using statistic methods. We apply our new method to study this important question.

6.1 *Data and implementation*

We let y_{it} be a measure of democracy for country i in period t and let X_{it} be the logarithm of its real GDP per capita. The measure of democracy and real GDP per capita are from the Freedom House and Penn World Tables, respectively.[7] Note that the Freedom House measures of democracy (y_{it}) are normalized to be between 0 and 1.

We consider the specification with both individual and time fixed effects,

$$y_{it} = \beta_{1i} X_{i,t-1} + \beta_{2i} y_{i,t-1} + \mu_i + \gamma_t + u_{it},$$

and assume that (β_{1i}, β_{2i}) has a group structure to account for possible heterogeneity. In a closely related paper, BM consider a group structure in the interactive fixed effects and assume (β_{1i}, β_{2i}) is constant for all is.[8]

We use a balanced panel data set similar to that in BM. The number of countries (N) is 74. The time index is $t = 1, \dots, 8$, where each period corresponds to a 5-year interval over the period 1961–2000. For example, $t = 1$ refers to years 1961–1965. Because the lagged y_{it} is used as a regressor, the number of time periods (T) is seven, that is, $t = 2, \dots, 8$. The choice of countries is determined by data availability. In addition, we exclude the countries whose measures of democracy remain constant over the seven periods $(t = 2, \dots, 8)$. A list of the 74 countries can be found in Table 8. Table 4 presents the summary statistics. The implementation details of our method are the same as in the simulations.

6.2 *Testing and estimation results*

We first test the hypothesis $\mathbb{H}_0(1)$, that is, we test whether (β_{1i}, β_{2i}) is constant for all i. We soundly reject this hypothesis with a p value being less than 0.001. This provides

[7]All the data are directly from AJRY: http://economics.mit.edu/faculty/acemoglu/data/ajry2008.

[8]In BM's Supplementary Material, they also consider the slope heterogeneity. Specifically, they consider the two specifications (using our notation)

$$y_{it} = \beta_{g_i}' X_{it} + \alpha_{g_i t} + u_{it},$$
$$y_{it} = \beta_{g_i}' X_{it} + \alpha_{g_i t} + \mu_i + u_{it},$$

where β_{g_i} is the group-specific slope coefficient, $\alpha_{g_i t}$ is the group-specific (time-varying) fixed effect, and μ_i is the country-specific (time-invariant) fixed effect. Their model specification is more general than ours, as they allow time-varying group-specific fixed effects $\alpha_{g_i,t}$. They report the results for the number of groups being four. For their first specification, they find that income effect is statistically significant, while for their second specification, they find that income effect is not statistically significant. Our empirical results are somehow different from theirs due to the different model specifications adopted.

TABLE 4. Summary statistics ($N = 74$).

Time Period t	Years	Y_{it}: Democracy			X_{it}: Logarithm of Real GDP per Capita		
		Mean	Median	s.d.	Mean	Median	s.d.
1	1961–1965	0.56	0.54	0.26	7.69	7.66	0.81
2	1966–1970	0.37	0.33	0.33	7.83	7.76	0.85
3	1971–1975	0.34	0.33	0.31	7.93	7.96	0.89
4	1976–1980	0.42	0.33	0.32	8.03	8.02	0.93
5	1981–1985	0.46	0.33	0.35	8.06	8.07	0.94
6	1986–1990	0.51	0.50	0.34	8.10	8.16	1.01
7	1991–1995	0.54	0.50	0.33	8.14	8.23	1.10
8	1996–2000	0.60	0.67	0.33	—	—	—

strong evidence that the slope coefficients are not homogeneous. We then test the hypothesis $\mathbb{H}_0(2)$ for three values of the tuning parameter $\lambda = c \cdot s_Y^2 \cdot T^{-1/3}$ ($c = 0.25$, 0.5, and 0.75). We reject this hypothesis at the 5% level with p values being 0.03, 0.03, and 0.02 for $c = 0.25$, 0.5, and 0.75, respectively. We continue to test $\mathbb{H}_0(3)$ and find that the p values are 0.42, 0.20, and 0.14 for $c = 0.25$, 0.5, and 0.75, respectively. Considering that the p values are above or close to the recommended nominal level $1/T$ (0.14), we stop the testing procedure and conclude that the number of groups is three. Table 5 presents all the testing results and Table 8 shows the country membership of the three groups. To take into account the issue of multiple testing, we also report Holm's (1979) adjusted p values in the last row in Table 5,[9] which also lend strong support to the conclusion of three groups in the data.

We also consider SSP's IC (see equation (5.1) above) to determine the number of groups. Again, we try three choices of λ as above with $c = 0.25$, 0.5, and 0.75. When $c = 0.25$ and 0.5, the IC determines three groups, as shown in Figure 1. When $c = 0.75$, the IC

TABLE 5. Test statistics.

Null Hypothesis	$c = 0.25$			$c = 0.5$			$c = 0.75$		
	$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$	$K = 3$
Statistics	3.57	1.93	0.20	3.57	1.93	0.83	3.57	2.06	1.07
p values	0.0002	0.03	0.42	0.0002	0.03	0.20	0.0002	0.02	0.14
Holm adjusted p values	0.0006	0.05	0.42	0.0006	0.05	0.20	0.0006	0.04	0.14

Note: Numbers in the main entries are the test statistics, p values, and Holm adjusted p values for three values of the constant c in the tuning parameter λ: 0.25, 0.5, and 0.75.

[9]Suppose that we have tested K^* individual hypotheses, and we order the individual p values from the smallest to the largest as $p_{(1)} \le p_{(2)} \le \cdots \le p_{(K^*)}$ with their corresponding null hypotheses labeled accordingly as $\mathbb{H}_{0(1)}, \mathbb{H}_{0(2)}, \ldots, \mathbb{H}_{0(K^*)}$. We calculate the step-down Holm adjusted p values for testing $\mathbb{H}_{0(k)}$ as

$$\text{adjusted } p_{(k)} = \min\bigl((K^* - k + 1) p_{(k)}, 1\bigr).$$

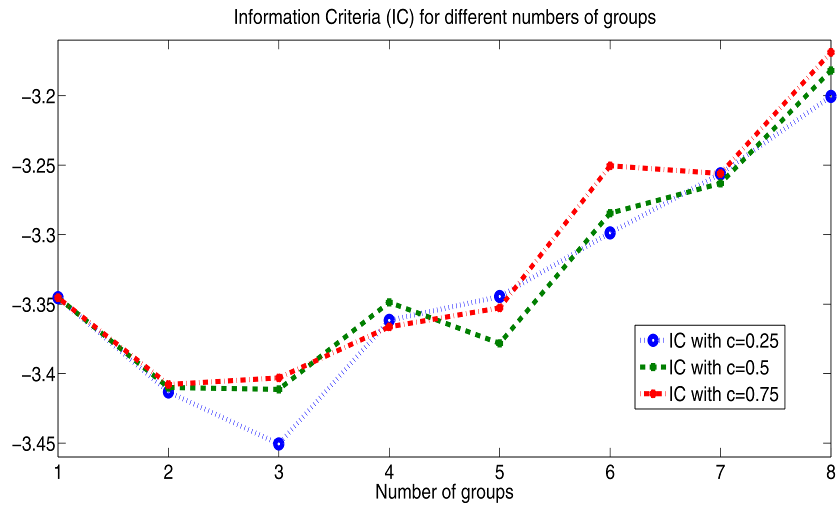

Figure 1. Information criteria. Values in the figure are the information criteria for different number of groups for three values of the constant c in the tuning parameter λ: 0.25, 0.5, and 0.75.

determines two groups.[10] Given that the majority of our results determine three groups, we conclude that $K = 3$.

Table 6 presents the estimation results. We report both C-Lasso estimates and post C-Lasso estimates. C-Lasso estimates are defined in Section 4.1. The post C-Lasso is implemented as in (4.5). Both estimates are bias-corrected and the standard errors are obtained using the asymptotic theory developed in SSP. The C-Lasso and post-C-Lasso estimation results are similar for different values of $c = 0.25$, 0.5, and 0.75. In the following discussion, we focus on the post-C-Lasso estimates with $c = 0.5$. It is clear that the estimated slope coefficients exhibit substantial heterogeneity. For β_{1i}, the estimates for the three groups are 0.26, −0.16, and −0.24. All of them are significant at the 1% level. It is interesting to note that not only the magnitudes, but also the signs of estimates, are different among the three groups. For Group 1, income has a positive effect on democracy, while for Groups 2 and 3, the effects are negative with different magnitudes. For β_{2i}, the three group estimates are 0.10, −0.24, and 0.41. The first estimate is not significant, while the second and third estimates are significant at the 1% level. We also present the point estimates of the cumulative income effect (CIE), which is defined as $\beta_{1i}/(1 - \beta_{2i})$. The estimates of CIE for the three groups are 0.28, −0.13, and −0.40, which imply that a 10% increase in income per capita is associated with increases of 0.028, −0.013, and −0.040 in the measures of democracy, respectively.

Note that if we assume that β_{1i} and β_{2i} are homogeneous across i, then the common estimates of β_{1i} and β_{2i} are −0.02 and 0.27 with t statistics being −0.45 and 5.95, respectively. This suggests that the effect of income on democracy is not statistically significant. This finding is consistent with that in AJRY. Nevertheless, the insignificance could be due to the model misspecification, which ignores slope heterogeneity. Note that the

[10]The values of IC for two groups and three groups are −3.4078 and −3.4031, respectively, which are actually quite close.

Table 6. Estimation results.

			β_{1i}			β_{2i}			CIE
			Estimates	s.e.	t Stat	Estimates	s.e.	t Stat	Estimates
Common Estimation			−0.02	0.038	−0.45	0.27	0.046	5.95	−0.02
$c = 0.25$	C-Lasso	Group 1	0.23	0.050	4.62	0.11	0.058	1.99	0.26
		Group 2	−0.21	0.032	−6.65	−0.31	0.055	−5.67	−0.16
		Group 3	−0.24	0.102	−2.36	0.53	0.052	10.28	−0.52
	Post-C-Lasso	Group 1	0.25	0.046	5.38	0.09	0.058	1.59	0.27
		Group 2	−0.19	0.033	−5.67	−0.24	0.041	−5.90	−0.15
		Group 3	−0.28	0.095	−2.97	0.49	0.049	9.99	−0.55
$c = 0.5$	C-Lasso	Group 1	0.27	0.042	6.38	0.11	0.065	1.71	0.30
		Group 2	−0.19	0.031	−5.98	−0.33	0.067	−4.91	−0.14
		Group 3	−0.21	0.084	−2.56	0.47	0.064	7.40	−0.40
	Post-C-Lasso	Group 1	0.26	0.045	5.66	0.10	0.064	1.51	0.28
		Group 2	−0.16	0.033	−4.80	−0.24	0.047	−5.05	−0.13
		Group 3	−0.24	0.078	−3.02	0.41	0.061	6.78	−0.40
$c = 0.75$	C-Lasso	Group 1	0.29	0.043	6.84	0.03	0.067	0.44	0.30
		Group 2	−0.19	0.032	−5.90	−0.29	0.056	−5.21	−0.14
		Group 3	−0.17	0.076	−2.25	0.49	0.058	8.52	−0.34
	Post-C-Lasso	Group 1	0.27	0.047	5.71	0.04	0.067	0.55	0.28
		Group 2	−0.16	0.033	−4.76	−0.24	0.047	−5.00	−0.13
		Group 3	−0.19	0.070	−2.72	0.43	0.054	8.06	−0.34

Note: Common estimation assumes one group. C-Lasso and post-C-Lasso estimates are based on three groups. We consider three values of the constant c in the tuning parameter λ: 0.25, 0.5 and 0.75. CIE stands for cumulative income effect, which is defined as $(\beta_{1i}/(1-\beta_{2i}))$.

common estimate falls in the middle of the group estimates. Here all group estimates are significant, but with different signs. Therefore, the common estimate, which can be thought of as a certain weighted average of the group estimates, could be close to zero and statistically insignificant.

To understand the heterogeneity in the data intuitively, we select three countries (Malaysia, Indonesia, and Nepal) and show their time-series data in Table 7. We simply calculate the correlations between the dependent variable Y_{it} and the key explanatory variable $X_{i,t-1}$. Even the simple correlations exhibit substantial heterogeneity with the values being −0.86, 0.07, and 0.66. This suggests that it is implausible for the slope coefficients to be the same for all countries, even without performing any sophisticated analysis.

This application shows that ignoring the heterogeneity in the slope coefficients can mask the true underlying relationship among economic variables.

6.3 *Explaining the group pattern*

As discussed above, the group estimates of β_{1i} and β_{2i} show substantial heterogeneity, though most of them are statistically significant at the 5% level. So far, our classification of the groups is completely statistical and does not use any a priori information. One

Table 7. Correlation between Y_{it} and $X_{i,t-1}$ for selected countries.

	Malaysia		Indonesia		Nepal	
Time Period t	Y_{it}	X_{it}	Y_{it}	X_{it}	Y_{it}	X_{it}
1	0.80	7.82	0.10	6.80	0.29	6.65
2	0.83	7.97	0.33	6.99	0.17	6.70
3	0.67	8.19	0.33	7.26	0.17	6.79
4	0.67	8.49	0.33	7.55	0.67	6.76
5	0.67	8.60	0.33	7.73	0.67	6.91
6	0.33	8.78	0.17	7.96	0.50	6.99
7	0.50	9.07	0	8.20	0.67	7.13
8	0.33	9.20	0.67	8.20	0.67	7.29
Correlation between Y_{it} and $X_{i,t-1}$	−0.86		0.07		0.66	

natural question is how to explain the group membership. Apparently, the group membership shown in Table 8 does not match the countries' geographic locations. We further investigate the group pattern using a cross-sectional multinomial logit model.

We let the dependent variable be group membership, which takes one of three values: 1, 2, or 3.[11] The explanatory variables include (i) a measure of constraints on the ex-

Table 8. Classification of countries.

Group 1 (Positive β_1 and Insignificant β_2) ($N_1 = 21$)				
Brazil	Cyprus	Algeria	Ecuador	Spain
Ghana	Greece	Jordan	Korea Rep.	Malawi
Nepal	Panama	Philippines	Portugal	Rwanda
Thailand	Taiwan	Uruguay	Venezuela RB	Congo Dem. Rep
Zambia				
Group 2 (Negative β_1 and Negative β_2) ($N_2 = 16$)				
Argentina	Congo Rep.	Dominican Republic	Finland	Gabon
Indonesia	Japan	Luxembourg	Mexico	Nigeria
Singapore	El Salvador	Togo	Tunisia	Turkey
Uganda				
Group 3 (Negative β_1 and Positive β_2) ($N_3 = 37$)				
Burundi	Benin	Burkina Faso	Bolivia	Central African Republic
Chile	China	Cameroon	Colombia	Egypt Arab Rep
Guinea	Guatemala	Guyana	Honduras	India
Iran	Israel	Jamaica	Kenya	Sri Lanka
Morocco	Madagascar	Mali	Mauritania	Malaysia
Niger	Nicaragua	Peru	Paraguay	Romania
Sierra Leone	Sweden	Syrian Arab Republic	Chad	Trinidad and Tobago
Tanzania	South Africa			

[11]We only report the results for $c = 0.5$. The results for $c = 0.25$ and 0.75 are similar and available upon request.

TABLE 9. Summary statistics by group.

		Group 1		Group 2		Group 3	
Variables	Variable Description	Mean	s.d.	Mean	s.d.	Mean	s.d.
Constraint	Constraints on the executive at independence	0.15	0.24	0.33	0.30	0.39	0.36
Growth	500-year income per capita change	2.00	1.18	2.26	1.09	1.50	0.95
democ	500-year democracy change	0.80	0.23	0.66	0.30	0.65	0.29
indcent	Year of independence/100	18.91	0.73	18.95	0.71	19.10	0.67
edu65	Education level in 1965	2.78	1.52	2.81	2.17	2.62	1.93
inc65	Logarithm of real GDP per capita in 1965	7.72	0.82	7.93	0.97	7.58	0.73
dem65	Measure of democracy in 1965	0.51	0.27	0.63	0.25	0.55	0.26

ecutive at independence, (ii) the 500-year change in income per capita over 1500–2000, (iii) the 500-year change in democracy over 1500–2000, (iv) independence year/100, (v) initial education level in 1965, (vi) initial income level in 1965, and (vii) initial democracy level in 1965. Acemoglu, Johnson, and Robinson (2005) argue that (i) is an important determinant of democracy. The variables (ii) and (iii) present long-run changes in income and democracy levels, respectively. The variable (iv) measures how recently a country became independent. Variables (v), (vi), and (vii) are the initial key economic variables. All the data are taken directly from AJRY.

Table 9 provides summary statistics for each of the three groups. The sample averages of (i) (the constraints on the executive at independence) are clearly substantially different among the three groups. For this variable, the average value of Group 1 is only about half that of Group 3. The variable of (ii) (the 500-year change in income per capita) differs noticeably among the three groups. Group 2 has the largest sample average of 2.26, while Group 3 has the smallest value of 1.50.

Table 10 presents the multinomial logit regression results for various model specifications. We choose Group 3 as the base group. Compared with Group 3, at the 5% level (a higher value of (i)), the constraints on the executive at independence lead to a reduced likelihood of being in Group 1. On the other hand, a higher value of (ii) (the 500-year change in income per capita) leads to a higher likelihood of being in Group 2. For the case $c = 0.5$ reported here, the variable (iv) (independence year/100) is also significant at the 5% level. However, this result is not robust to other choices of c. In summary, we find that the constraints on the executive at independence and the long-run economic growth are important determinants of our group pattern.

7. CONCLUSION

We develop a data-driven testing procedure to determine the number of groups in a latent group panel structure proposed in SSP. The procedure is based on conducting hypothesis testing on the model specifications. The test is a residual-based LM-type test and is asymptotically normally distributed under the null. We apply our new method to study the relationship between income and democracy, and find strong evidence that

Table 10. Determinants of the group pattern.

Group 1							
Constraints	−3.01**	−3.26***	−3.04***	−5.20***	−5.55***	−5.92***	−5.99***
	(1.20)	(1.18)	(1.13)	(1.53)	(1.68)	(1.81)	(1.89)
Growth		0.60**	0.58*	0.96**	1.16*	1.67**	1.69**
		(0.29)	(0.35)	(0.41)	(0.61)	(0.72)	(0.74)
demco			1.65	3.94**	4.95*	4.69*	4.48*
			(1.40)	(1.85)	(2.66)	(2.70)	(2.66)
indcent				1.77**	2.12**	1.92**	1.97**
				(0.77)	(0.89)	(0.94)	(0.96)
edu65					−0.34	−0.14	−0.08
					(0.34)	(0.37)	(0.41)
inc65						−1.41	−1.36
						(0.93)	(0.96)
dem65							−0.53
							(2.02)
Group 2							
Constraints	−0.50	−1.03	−0.99	−1.39	−1.58	−1.99	−2.44
	(0.91)	(0.95)	(0.98)	(1.29)	(1.62)	(1.74)	(1.83)
Growth		0.72**	0.89**	1.02**	1.19**	1.76**	1.92**
		(0.32)	(0.37)	(0.41)	(0.60)	(0.77)	(0.80)
demco			−0.98	−0.71	−0.67	−0.73	−0.56
			(1.22)	(1.42)	(2.27)	(2.39)	(2.32)
indcent				0.35	0.24	0.06	0.10
				(0.70)	(0.92)	(0.99)	(1.00)
edu65					−0.29	−0.09	−0.25
					(0.35)	(0.38)	(0.42)
inc65						−1.40	−1.62
						(0.99)	(1.05)
dem65							1.97
							(2.17)

The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The results are based on a multinomial logit regression where Group 3 is taken as the reference group. The dependent variable is the group membership. The standard errors are calculated without taking into account the fact that the dependent variables are estimated.

the slope coefficients are heterogeneous and form three distinct groups. Further, we find that the constraints on the executive at independence and long-run economic growth are important determinants of the group pattern.

There are several interesting topics for further research. Here we apply our testing procedure to determine the number of groups for slope coefficients. The same idea can be applied to other group structures, such as those considered in BM, where fixed effects have a grouped pattern. We may also extend our methods to nonlinear panel data models such as discrete choice models.

References

Acemoglu, D., S. Johnson, and J. Robinson (2005), "The rise of Europe: Atlantic trade, institutional change, and economic growth." *American Economic Review*, 95, 546–579.

[757]

Acemoglu, D., S. Johnson, J. Robinson, and P. Yared (2008), "Income and democracy." *American Economic Review*, 98, 808–842. [731]

Ando, T. and J. Bai (2016), "Panel data models with grouped factor structure under unknown group membership." *Journal of Applied Econometrics*, 31, 163–191. [729, 730, 734, 737, 738]

Barro, R. J. (1999), "Determinants of democracy." *Journal of Political Economy*, 107, 158–183. [751]

Bester, C. A. and C. B. Hansen (2016), "Grouped effects estimators in fixed effects models." *Journal of Econometrics*, 190, 197–208. [729, 730]

Bonhomme, S. and E. Manresa (2015), "Grouped patterns of heterogeneity in panel data." *Econometrica*, 83, 1147–1184. [729]

Deb, P. and P. K. Trivedi (2013), "Finite mixture for panels with fixed effects." *Journal of Econometric Methods*, 2, 35–51. [729, 730]

Hahn, J. and G. Kuersteiner (2011), "Reduction for dynamic nonlinear panel models with fixed effects." *Econometric Theory*, 27, 1152–1191. [737]

Hahn, J. and H. R. Moon (2010), "Panel data models with finite number of multiple equilibria." *Econometric Theory*, 26, 863–881. [730]

Holm, S. (1979), "A simple sequentially rejective multiple test procedure." *Scandinavian Journal of Statistics*, 6, 65–70. [753]

Hsiao, C. (2014), *Analysis of Panel Data*. Cambridge University Press, Cambridge. [729, 741]

Hsiao, C. and A. K. Tahmiscioglu (1997), "A panel analysis of liquidity constraints and firm investment." *Journal of the American Statistical Association*, 92, 455–465. [732]

Leeb, H. and P. M. Pötscher (2005), "Model selection and inference: Facts and fiction." *Econometric Theory*, 21, 21–59. [740]

Leeb, H. and P. M. Pötscher (2008), "Sparse estimators and the oracle property, or the return of Hodges' estimator." *Journal of Econometrics*, 142, 201–211. [740]

Leeb, H. and P. M. Pötscher (2009), "On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding." *Journal of Multivariate Analysis*, 100, 2065–2082. [740]

Lin, C.-C. and S. Ng (2012), "Estimation of panel data models with parameter heterogeneity when group membership is unknown." *Journal of Econometric Methods*, 1, 42–55. [729, 730, 734]

Lipset, S. M. (1959), "Some social requisites of democracy: Economic development and political legitimacy." *American Political Science Review*, 53, 69–105. [751]

Onatski, A. (2009), "Testing hypotheses about the number of factors in large factor models." *Econometrica*, 77, 1447–1479. [730, 733]

Pesaran, M. H., R. Smith, and K. S. Im (1996), "Dynamic linear models for heterogeneous panels." In *The Econometrics of Panel Data: A Handbook of the Theory With Applications*, second edition (L. Matyas and P. Sevestre, eds.), 145–195, Kluwer Academic Publishers, Dordrecht. [731]

Pesaran, M. H. and T. Yamagata (2008), "Testing slope homogeneity in large panels." *Journal of Econometrics*, 142, 50–93. [731, 738]

Phillips, P. C. B. and D. Sul (2003), "Dynamic panel estimation and homogeneity testing under cross section dependence." *Econometrics Journal*, 6, 217–259. [731]

Sarafidis, V. and N. Weber (2015), "A partially heterogeneous framework for analyzing panel data." *Oxford Bulletin of Economics and Statistics*, 77, 274–296. [729, 730, 734]

Schneider, U. and P. M. Pötscher (2009), "On the distribution of the adaptive LASSO estimator." *Journal of Statistical Planning and Inference*, 139, 2775–2790. [740]

Su, L. and Q. Chen (2013), "Testing homogeneity in panel data models with interactive fixed effects." *Econometric Theory*, 29, 1079–1135. [731, 737]

Su, L. and G. Ju (forthcoming), "Identifying latent grouped patterns in panel data models with interactive fixed effects." *Journal of Econometrics*. [729]

Su, L., Z. Shi, and P. C. B. Phillips (2016), "Identifying latent structures in panel." *Econometrica*, 84, 2215–2264. [729]

Sun, Y. (2005), "Estimation and inference in panel structure models." Working Paper, Department of Economics, UCSD. [729, 730]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 24 November, 2014; final version accepted 27 January, 2017; available online 2 February, 2017.